

NASA Contractor Report 185637

National Aeronautics and Space Administration (NASA)/
American Society for Engineering Education (ASEE)
Summer Faculty Fellowship Program—1990

Volume 2

Richard B. Bannerot, Editor
University of Houston — University Park
Houston, Texas

Stanley H. Goldstein, Editor
University Programs Office
Lyndon B. Johnson Space Center
Houston, Texas

Grant NGT 44-005-803
December 1990

NASA

(NASA-CR-185637-Vol-2) NASA/ASEE SUMMER
FACULTY FELLOWSHIP PROGRAM, 1990, VOLUME 2
(Houston Univ.) 177 p CSCL 058

N91-27103
--THRU--
N91-27117
Unclas

63/99 0020215

PREFACE

The 1990 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by the University of Houston—University Park and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are

1. To further the professional knowledge of qualified engineering and science faculty members
2. To stimulate an exchange of ideas between participants and NASA
3. To enrich and refresh the research and teaching activities of participant's institutions
4. To contribute to the research objectives of the NASA Centers

Each faculty fellow spent at least 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1990. Volume 1 contains reports 1 through 14, and volume 2 contains reports 15 through 28.

CONTENTS

Volume 1

1.	Akundi, Murty A.: "LIF and Emission Studies of Copper and Nitrogen"	1-1
2.	Barnes, Ron: "Compound Estimation Procedures in Reliability"	2-1
3.	Bishop, Phillip A.: "Techniques for Determining Total Body Water Using Deuterium Oxide"	3-1
4.	Boyd, Ernest J.: "Mathematical Modeling of the Flow Field and Particle Motion in a Rotating Bioreactor at Unit Gravity and Microgravity"	4-1
5.	Hart, Maxwell M.: "An Air Revitalization Model for Regenerative Life Support Systems (RLSS)"	5-1
6.	Hite, Gerald E.: "Photographic Image Enhancement"	6-1
7.	Hooker, John C.: "Nonlinear Systems Dynamics in Cardiovascular Physiology: The Heart Rate Delay Map and Lower Body Negative Pressure"	7-1
8.	Huntsberger, Terry: "Incorporation of Shuttle CCT Parameters in Computer Simulation Models"	8-1
9.	Johnson, Debra Steele: "Training Effectiveness of an Intelligent Tutoring System for a Propulsion Console Trainer"	9-1
10.	Karimi, Amir: "Re-Examination of Metman, Recommendations on Enhancement of LCVG, and Development of New Concepts for EMU Heat Sink"	10-1
11.	Kurdila, Andrew J.: "A Nonrecursive Order N Preconditioned Conjugate Gradient/Range Space Formulation of MDOF Dynamics"	11-1
12.	Lacovara, R. C.: "Applications of Formal Simulation Languages in the Control and Monitoring Subsystem of Space Station Freedom"	12-1
13.	Lawless, DeSales: "A Study of Murine Bone Marrow Cells Cultured in Bioreactors Which Create an Environment Which Simulates Microgravity"	13-1
14.	Layne, Charles S.: "The Effects of Deuterium on Static Posture Control"	14-1

Volume 2

15.	Morgan, Paul: "Evaluation of Geophysical Properties of the Lunar Regolith for the Design of Precursor Scientific Missions for the Space Exploration Initiative"	15-1
16.	Morris, Robert A.: "A Design for an Intelligent Monitor and Controller for Space Station Electrical Power Using Parallel Distributed Problem Solving"	16-1

17.	Murphy, Michael G.: "Fuzzy Logic Control System to Provide Autonomous Collision Avoidance for Mars Rover Vehicle"	17-1
18.	Navard, Sharon E.: "Evaluating the Effect of Accuracy Ratios on the Percent of Calibrations Which Are Out of Tolerance"	18-1
19.	Neubek, Deborah J.: "Developing a Taxonomy for Mission Architecture Definition"	19-1
20.	Radhakrishnan, Ramalingam: "Design of a Lunar Oxygen Production Plant"	20-1
21.	Rajan, Periasamy K.: "Design of Correlation Filters for Hybrid Vision Systems"	21-1
22.	Ricles, James M.: "Development of Load-Dependent Ritz Vector Method for Structural Dynamic Analysis of Large Space Structures"	22-1
23.	Squires, William G.: "Mechanics, Impact Loads and EMG on the Space Shuttle Treadmill"	23-1
24.	Strait, Melissa M.: "Solutions to Problems of Weathering in Antarctic Eucrites"	24-1
25.	Tietze, Karen J.: "Assessment of the Pharmacodynamics of Intranasal, Intravenous and Oral Scopolamine"	25-1
26.	Uhde-Lacovara, J.: "High Accuracy Optical Rate Sensor"	26-1
27.	Williams, Raymond: "Planetary Spacecraft Cost Modeling Utilizing Labor Estimating Relationships"	27-1
28.	Yin, Paul K.: "Structural Concept Studies for a Horizontal Cylindrical Lunar Habitat and a Lunar Guyed Tower"	28-1

5-91
N 9 1 - 2 7 1 0 4 3

EVALUATION OF GEOPHYSICAL PROPERTIES OF THE LUNAR
REGOLITH FOR THE DESIGN OF PRECURSOR SCIENTIFIC MISSIONS
FOR THE SPACE EXPLORATION INITIATIVE

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

U 585 1445

Prepared By: Paul Morgan, Ph.D.
Academic Rank: Professor
University & Department: Northern Arizona University
Geology Department, Box 6030
Flagstaff, Arizona 86011-6030

NASA/JSC
Directorate: Space and Life Sciences
Division: Solar System Exploration
Branch: Space Resources Utilization Office
JSC Colleague: Wendell W. Mendell, Ph.D.
Date Submitted: August 7, 1990
Contract Number: NGT-44-005-803

ABSTRACT

A working knowledge of the physical parameters of the upper ten meters of the lunar regolith is crucial to the planning of experimentation, mining, resource utilization, and construction associated with future lunar bases. Scientific research literature pertaining to studies of the lunar surface, including results of Apollo experiments and exploration, have been reviewed to produce a summary of our present knowledge of the physical properties of the lunar regolith. Particular emphasis was placed on gaps in our understanding. The primary data source for this review has been Carrier *et al.* (1).

The purpose of this review is to address the following questions: 1) With what frequency may boulders be encountered that represent hazards to lunar operations?; 2) How easy may lunar soil be excavated?; 3) How may explosives be used in excavation operations?; 4) What is the trafficability of the regolith?; 5) What problems may be encountered in mining (probably strip mining) of the regolith?; 6) What angle(s) of repose will be stable in excavation of the regolith?; 7) What layering may be encountered in the subsurface?; 8) Is our knowledge of the regolith site specific, or may it be applied generically to any site on the lunar surface?; 9) What data will be needed to characterize a site for a lunar base?; 10) How may the regolith properties modify the design of geophysical experiments from the lunar base?; 11) Are there terrestrial analogues for the geophysical properties of the lunar regolith?

Available data indicate that there is significant lateral heterogeneity in the lunar regolith. This heterogeneity exists both with respect to small-size particle distributions (< 1 mm), which may affect bulk material properties, and with respect to the random distribution of large-sized particles (> 100 mm), which need to be considered for mining and excavation operations. Vertical layering of the regolith is expected at most sites, and considerable vertical heterogeneity may also be expected. Detailed site-specific surveys are recommended at all sites where mining and excavation operations are planned. These surveys should include sub-surface profiling, for which radar techniques appear to be most easily adapted, to determine the density and distribution of large, consolidated particles. This profiling should also determine the degree of subsurface layering. Core samples will be required to determine small-sized particle properties if significant layering is recorded. Trafficability appears to be primarily restricted by surface topographic features, although problems of raising the loose surface regolith must be considered. These problems appear to be associated with ballistic trajectories associated with surface activities which disturb the surface, and with electrostatic attraction by the lighter particles to charged surfaces. No terrestrial analogue adequately covers the range of properties likely to be experienced in the lunar regolith, but analogues for different physical and mechanical properties may be found.

INTRODUCTION

The lunar regolith is an essentially continuous layer, typically several meters in thickness, which covers the entire lunar surface (2). Any surface activities on the Moon, and any geological studies of the Moon are therefore strongly affected by the properties of the regolith. It is a mantle of fragmental and unconsolidated rock material of highly varied character, ranging from very fine dust to blocks of several meters in diameter. All lunar landings (Apollo and Luna), and all photographic studies of the Moon indicate that the regolith completely covers the underlying bedrock, except perhaps where bedrock may be exposed on some very steep-sided crater walls and lava channels. The finer components of the lunar regolith (<10 mm) are often informally referred to as lunar "soil", but this term must be used carefully as there are fundamental differences between terrestrial regoliths, of which soils are one group, and the lunar regolith (see DIFFERENCES AMONG THE MODES OF FORMATION OF LUNAR AND TERRESTRIAL REGOLITHS below).

Basically the constitution of the lunar regolith is relatively simple in comparison with terrestrial regoliths. There are five basic particle types in the lunar regolith:

- 1) Crystalline rocks (fragments of multi-mineral rocks which formed by crystallization from silicate melt)
- 2) Minerals (fragments of individual mineral crystals)
- 3) Breccias (fragments of rocks which are themselves composed of an aggregate of other rock and mineral fragments)
- 4) Agglutinates (pulverized rock, mineral and glass fragments bonded together with glass)
- 5) Glasses (fragments, irregular masses and beads of amorphous silicates formed by rapid cooling of silicate melts)

Fewer than a hundred minerals have been identified in lunar samples, compared with several thousand minerals that have been identified on Earth. While we may expect new minerals to be identified with continuing lunar exploration, we may confidently extrapolate present knowledge of lunar mineralogy to indicate that the mineralogy of lunar crust is at least an order of magnitude simpler than terrestrial crustal mineralogy. The physical properties of the lunar regolith also fall into relatively limited ranges compared to those of terrestrial regoliths due to the relatively simple mineralogy, the absence of a significant atmosphere, and the lack of water, clay minerals, and organic material in the lunar regolith. The primary factors that affect the physical properties of the lunar regolith are the packing of the regolith (bulk and relative densities) and temperature.

Current scenarios for the establishment of a permanent manned base on the Moon require specific activities which interact with the regolith (3). These activities include: 1) Use of regolith material to shield long-term working and habitation modules sufficiently to reduce radiation from the largest recorded solar flares to safe

levels; 2) Surface transportation across the regolith; 3) Excavation of regolith material for oxygen production, and in the future for other mining operations such as the production of rare isotopes, such as ^3He , implanted in the regolith by the solar wind. A number of specific questions need to be answered before the mechanics of these activities can be evaluated. These questions include: 1) With what frequency may boulders be encountered that represent hazards to lunar operations?; 2) How easy may lunar soil be excavated?; 3) How may explosives be used in excavation operations?; 4) What is the trafficability of the regolith?; 5) What problems may be encountered in mining (probably strip mining) of the regolith?; 6) What angle(s) of repose will be stable in excavation of the regolith?; 7) What layering may be encountered in the subsurface?; 8) Is our knowledge of the regolith site specific, or may it be applied generically to any site on the lunar surface?; 9) What data will be needed to characterize a site for a lunar base?; 10) How may the regolith properties modify the design of geophysical experiments from the lunar base?; 11) Are there terrestrial analogues for the geophysical properties of the lunar regolith? Physical properties of the lunar regolith have been measured in situ by robots and astronauts, in the laboratory on returned samples, and by remote sensing from the Earth's surface and from lunar orbit. This report summarizes these results as they apply to the establishment of a permanent manned lunar base, and uses these results to propose precursor scientific studies for the Space Exploration Initiative. For the purposes of this report, the discussion of physical properties has been limited to mechanical (geotechnical) properties, which include parameters which describe grain size and shape, densities, and other geotechnical properties. These data are primarily summarized from Carrier *et al.* (1). Other physical properties, which include elastic (seismic), electrical and electromagnetic, magnetic, and thermal properties will be considered in future work. The section on mechanical properties is followed by a discussion of the implications of these properties, the mode of formation of the regolith, and considerations for lunar base activities.

MECHANICAL (GEOTECHNICAL) PROPERTIES

The mechanical properties of the lunar regolith were of vital importance to design of the Apollo Lunar Module and there was significant concern regarding the implications of these properties until the first human footsteps were made on the Moon (1). Some of the major uncertainties resulted from lack of knowledge of the behavior of fragmental lunar material in the lunar vacuum. Predictions varied from a very hard surface due to "cold welding" of the fragments in the vacuum, to a very soft surface due to the very small grain size of the fragments. Our present experience is that the lunar surface is not very soft, nor very hard. The range of geotechnical properties of the lunar regolith are less than occur in the terrestrial regolith: There appear to be three sources of this difference: 1) Terrestrial geologic exogenic processes tend to produce well-sorted sediments; The primary lunar soil-forming process is meteorite impact, which produces a poorly-sorted regolith;

2) Geotechnical properties of terrestrial soils vary by orders of magnitude as a result of water, clay minerals, and organic materials, not present on Moon; 3) The range of minerals in lunar soils much less than on Earth, in fact a large portion of the soil is glass.

Geotechnical properties of lunar soil tend to fall in a fairly narrow range, and the most significant variable is the relative density. Most other physical properties of the lunar regolith also tend to fall in a relatively narrow range, the largest variations occurring in those properties dependant on the relative density.

Particle Size Distribution

The particle size distribution of a regolith controls to various degrees its strength, compressibility, optical, thermal, and elastic (seismic) properties. The size distribution of selected samples have been summarized by Carrier et al. (1). Care should be exercised in using these data as they were derived from biased samples in the sense that only the soil fraction of the regolith, fragments smaller than 1 cm, have been included in the particle size distribution analysis. The consequences of this biased sampling are discussed below.

Carrier (4) has interpreted these results to indicate that the bulk of lunar soil samples fall in fairly narrow range of particle size distribution. In general, these results indicate that the soil is a poorly sorted, silty sand to sandy silt. The median particle size in the <1 mm fraction is 40 to 130 microns, with an average of 70 microns. In practical terms this means that about half of the lunar soil by weight is finer than human eye can resolve. Roughly 10-20% of soil finer than 20 microns, and a thin layer of dust appears to adhere electrostatically to everything that comes in contact with the soil: spacesuits, tools, equipments, lenses. Carrier *et al.* (1) warn that **"Housekeeping is a major challenge for operations on the lunar surface."**

Regolith dust is clearly a factor which must be taken into account in lunar surface operations, but potential problems with large regolith fragments is not so clearly illustrated by these data. Even some of the samples listed by Carrier et al. (1) have a significant fraction of particles with grain sizes greater than 1 cm. The bulk of samples 15400 and 70180 consist of these large particles, which were excluded from the median grain size calculations. The magnitude of this problem is clear from an examination of almost any of the photographs taken on the lunar surface (e.g., (5), Figures 4.1 and 4.3). Many centimeter- to meter-sized fragments are visible in these photographs, and in places large fragments clearly constitute a significant portion of the regolith. There is a need for greater study of the size and frequency distribution of the larger regolith particles.

Particle Shapes

Shapes of the lunar regolith particles are highly variable, ranging from spherical to extremely angular. In general, particles are somewhat elongated and are

sub-angular to angular. Because of this elongation, particles tend to pack together with a preferred orientation of the long axes, and physical properties in situ are expected to be anisotropic. Particles are not compact, but have irregular, re-entrant surfaces, a feature which especially affects the shear strength of the regolith.

Studies of the average shape properties of the finer fraction of the regolith indicate that the particles have an average elongation of about 1.35 (ratio of the major and intermediate axes of the fragments), and aspect ratios for most of the particles are in the range of 0.4 to 0.7 (ratio of minor and major axes of an ellipse fitted to the particle shape by least squares). Most particles are therefore described as slightly-to-medium elongated. Measurements of particle roundness indicate that the particles are sub-angular to angular in shape.

Specific Gravity

The specific gravity of a soil particle is defined as the ratio of its mass to the mass of an equal volume of water at 4°C. Values for lunar soils range from 2.3 to >3.2 and Carrier *et al.* (1) recommend a representative value of 3.1 for general scientific and engineering purposes. In contrast, many terrestrial soils have a specific gravity of 2.7. The average specific gravity of a given lunar soil is related to the relative proportions of the different particle types which constitute the soil, i.e., the relative proportions of basalts, mineral fragments, breccias, agglutinates, and glasses.

Bulk Density and Porosity

The bulk density of a soil is defined as the mass of the material contained in unit volume (usually given in g cm^{-3} or Mg m^{-3}). Porosity is defined as the volume of void space in the material divided by its total volume. The *in situ* bulk density of a lunar soil is a fundamental property, influencing its bearing capacity, slope stability, seismic velocity, thermal conductivity, electrical resistivity, and the depth of penetration of ionizing radiation. Lunar regolith bulk density is highly variable, not only from site to site, but also from station to station, and even with depth in a single core. At present, the best estimate for the average bulk density of the top 150 mm of lunar soil is $1.50 \pm 0.05 \text{ Mg m}^{-3}$, and of the top 600 mm, $1.66 \pm 0.05 \text{ Mg m}^{-3}$ (1). The average porosity in the top 300 mm of the surface is 49%. Bulk densities tend to be less variable below about 1 m, and the mean bulk density of a long core from Apollo 17 from a depth of 1 to 2 m is approximately 1.8 Mg m^{-3} . A variability of $\pm 0.15 \text{ Mg m}^{-3}$ is shown in bulk density measurements on cores from Apollo 15, 16 and 17 at these depths (1), however, and this variability could be significant in lunar base engineering activities.

Relative Density

Relative density is a measure of the degree of particle packing of a granular

soil, such as lunar soil. In situ relative density is another parameter which influences the physical and geophysical properties of a soil, and some of these properties may vary over several orders of magnitude between relative density values of 0% to 100%. *In situ* relative density of lunar soil has been found to be about 65% (medium to dense) in the top 150 mm, increasing to more than 90% (very dense) below a depth of 300 mm. These relative densities are very high when compared to a relative density of 65-75% which is the practical limit for field compaction of terrestrial soils. This high relative density of lunar soils was probably caused by the shaking effects of shock waves generated by innumerable meteoroid impacts.

Compressibility

Compressibility describes the volume change that occurs when a confining stress is applied to soil. It is quantified by the compression index, the decrease in void ratio that occurs when the stress is increased by an order of magnitude. The compression of lunar soils, measured in a vacuum, has been found to vary from 0.01 to 0.11, depending on the initial density and the applied stress. The compressibility curves for soils of different initial densities converge when the stress exceeds 100 kPa, which corresponds, on the Moon, to a depth of more than 30 m below the lunar surface.

Shear Strength

The shear strength consists of two components, a cohesive component, that is independent of the applied stress, and a frictional component that is directly proportional to the normal stress. Based upon a variety of data the following estimates for these components have been derived:

Cohesion = 0.1 to 1 kPa

Friction angle = 30° to 50°

The higher shear-strength parameters correspond to higher values of relative density of the lunar soil. These values are relatively high, probably due to the interlocking of irregular and re-entrant soil particles under the low stresses present near the lunar surface. These values are thought to be useful for near-surface lunar structure and construction activities, but for heavy structures or deep excavations, a more complete study of lunar soil shear strength should be made.

Permeability and Diffusivity

The coefficients of permeability and diffusivity are used to evaluate the movement of fluid through a porous medium, and no direct measurements of these parameters have been made on lunar soil. Based on a test firing of the Surveyor 5 vernier engine on the lunar surface, the permeability of the soil to a depth of 250 mm was estimated to be 1 to 7 x 10⁻¹² m². Diffusion rates for noble gases

through lunar soil have been measured in the laboratory yielding values of 7700, 2300, and 1800 mm² s⁻¹ for He, Ar, and Kr, respectively.

Bearing Capacity, Slope Stability and Trafficability

These three parameters control important engineering properties of the regolith. The ultimate bearing capacity defines the maximum possible load that can be applied without causing gross failure, and has been determined to be approximately 6,000 kPa for a one-meter footing on the lunar surface. The allowable bearing capacity defines a lesser load that can be applied without exceeding a given amount of settlement. For a maximum settlement of 10 mm for a load with a footing of less than 500 mm, data from the depths of astronaut boot-prints indicate, that for a 95% confidence level, the allowable bearing capacity should not exceed 2 kPa.

The slope stability of lunar soil is expected to vary widely from naturally compacted material to excavated material. Calculations indicate that a vertical cut to 3 m could be made in naturally compacted lunar soil, and a slope of 60° could be maintained to a depth of about 10 m. In contrast, a pile of uncompacted excavated soil could be raised to height of 10 m at an angle of nearly 40°.

Trafficability is defined as the capacity of a soil to support a vehicle and to provide sufficient traction for movement. From the experience of Apollo and Lunokhod missions, we know that almost any vehicle with round wheels will perform satisfactorily on the lunar surface provided that the ground contact pressure is no greater than about 7 to 10 kPa. The primary limitations on the trafficability of the lunar soil are speed and slope. The steepest slope that the Apollo Lunar Roving Vehicle could climb was 19° to 23°. Additional slope climbing capability could be attained by adding more aggressive tires to the vehicle. Particularly soft soil conditions were encountered by both Apollo and Lunokhod vehicles when approaching craters, particularly on the inside wall. Local problem areas may therefore exist for trafficability.

DIFFERENCES AMONG THE MODES OF FORMATION OF TERRESTRIAL AND LUNAR REGOLITHS

In order to understand the differences in physical properties among lunar and terrestrial regoliths one must first understand the basic differences in the modes of formation of these materials. Both regoliths are ultimately formed by breakdown of bedrock. Terrestrial regoliths are formed by a range of processes which commonly involve physical disintegration and some chemical change of the bedrock, transportation of the sediment so formed by wind, water or ice, commonly with additional physical and/or chemical modification, and redeposition of the sediment. These processes typically result in the production of a sediment which is relatively well-sorted both physically and chemically, and which is characterized by rounded particles. Sediments generally become smaller in grain size, more rounded, and more

chemically mature (stable in a hydrous environment at standard temperatures and pressures) as the distance of transportation from the bedrock source increases. The major exception to this generalization is erosion and direct deposition from ice which produces very poorly sorted sediments (grain sizes from rock flour to boulders), with little or no chemical change from the parent bedrock. The ice-transported sediments are rounded, however, and if ice melting occurs before sediment deposition, the sediments sizes are significantly sorted by transportation in the melt-waters. The sediments may then be further modified by soil formation, which causes further vertical physical and chemical differentiation and modification of the sediments and the addition of organic material to the upper regolith. The result is a regolith which is well-sorted with a primary vertical stratification typically modified by lateral facies changes.

In contrast, the lunar regolith is primarily the result of a single physical process, hypervelocity impacts (projectiles with impact velocities in excess of the speed of sound in the target material) of particles ranging in size from dust to small planets. These impacts cause major disintegration of the target material, typically forming angular fragments ranging in size from rock flour to large boulders, and transportation of these materials from the impact site is limited to ballistic projection of the fragments. Ballistic fragments with an initial velocity of greater than 2.4 km s^{-1} may be ejected from the Moon; particles with lower velocities are redeposited on the lunar surface, the larger particles forming *secondary craters*.

There is no sorting of particle size during ballistic transportation on the Moon as there is no significant atmosphere to cause differential hydrodynamic drag on particles of different sizes and shapes, as occurs on Earth. Chemical changes are limited to phase-transformations by high shock-pressures produced by impact, and changes in mineralogy associated with impact melting. These processes results in a suite of characteristic high-pressure minerals found only associated with impact structures. Impact melting results in a high proportion of rounded beads of glass in the lunar regolith, formed by ballistically projects melt droplets which rapidly solidify during flight to form amorphous glass beads before landing.

Although the result of repeated impacts is to continue to decrease grain size, reworking of the lunar regolith has not in general proceeded to a state of dynamic equilibrium in which the decrease in grain size due to impact fracturing is equal to the increase in grain size associated with melt solidification. The lunar regolith therefore consists of a poorly sorted aggregate of angular brecciated fragments and glass spherules. Layering in the regolith is associated with the overlap of ejecta blankets from different impact targets, and this layering tends to be modified by subsequent impact events (*gardening*). There is no lunar soil-forming processes equivalent to the hydro-chemical and organic processes of terrestrial soil formation. However, there is surficial regolith modification by exposure to the solar wind and micro-meteorite impacts.

DISCUSSION OF THE IMPLICATIONS OF REGOLITH PHYSICAL PROPERTIES FOR LUNAR BASE ACTIVITIES

From this preliminary review of the physical properties of the lunar regolith, I have identified two areas of potential concern for lunar base activities. These areas concern the probability of encountering cobbles and boulder in the regolith subsurface, and "dust" and heterogeneity in the fine fraction of the regolith.

Most of the published discussions on particle-size distributions are based upon data only from sieved samples of fines (< 1 mm) or the coarse and fine fraction (< 10 mm), ignoring the "lumps" (> 10 mm) in the regolith. Coarse fragments of the regolith samples were hand-picked from the samples in the Receiving Laboratory and given their own numbers - one may therefore be able to reconstruct the original sample particle-size constitution, but apparently these data are not readily available at present. There is also the problem that meter-sized boulders were not sampled *en masse* for obvious practical reasons, so this particle size is completely missing from the sample analyses. Photo-geologic studies have been made of the larger boulder distributions, and more work is needed on this topic. In a conservative approach, we may not ignore the possibility of excavating large boulders even in the mare regolith distant from large, young impacts. Cobbles and boulders are visible on many of the lunar surface photographs, and if similar conditions exist in the subsurface, this large-size component of the regolith could cause engineering hazards.

Problems of "dust" from the lunar regolith have been briefly discussed by a number of authors (e.g., (1)). There appear to be inconsistencies in some of these reports, however. Although there are many references to "dust permeating the samples" and "how quickly things were made dirty" in the astronaut voice communications of the Apollo lunar surface activities, the rock samples that were returned to the Lunar Receiving Laboratory at the Johnson Space Center were considered to be covered in dust only to the extent that would be expected from the packing of friable samples and transportation back to Earth (Don Morrison, personal communication, 1990). Some appearance of dusty-surfaces could have been the effects of darkening by exposure. The magnitude of the problems of dust associated with lunar surface activities remains unresolved at present. The main dust problems associated with lunar surface activities appear to be associated with porous surfaces, such as the space suits, or charged surfaces, camera lenses and other insulators. Other materials may still be mantled in dust by activities which cause ballistic disturbances of the regolith, but there is no evidence dust jumps onto all exposed surfaces. This dust may still be a major problem, and more study of this problem is required before extensive lunar surface activities are planned.

I am also concerned that not enough attention has been given to heterogeneity in particle-size in the fines. There is considerable variability in the mean grain size of the < 1 mm fraction (41-123 microns), and I suspect that this variation may result in variations in physical properties that are being homogenized in the averages which are quoted. For example, I suspect that the significance of re-entrant surfaces

becomes less as the particle size decreases, which could have measurable effects upon soil cohesiveness and related properties. This topic should be studied in greater detail.

RECOMMENDATION FOR SITE-SURVEYS

Available data indicate that there is significant lateral heterogeneity in the lunar regolith. This heterogeneity exists both with respect to small-size particle distributions (<1 mm), which may affect bulk material properties, and with respect to the random distribution of large-sized particles (>100 mm), which need to be considered for mining and excavation operations. Vertical layering of the regolith is expected at most sites, and considerable vertical heterogeneity may also be expected. Detailed site-specific surveys are recommended at all sites where mining and excavation operations are planned. These surveys should include sub-surface profiling to determine the density and distribution of large, consolidated particles. This profiling should also determine the degree of subsurface layering. Core samples will be required to determine small-sized particle properties if significant layering is recorded.

These surveys would be desirable, and perhaps essential for engineering purposes if the boulder problem is thought to be significant. They would be desirable for geophysical studies under any circumstances, primarily because the rapid change in physical properties of the upper regolith with depth (increase in compaction, etc.). Lateral variations in regolith thickness may cause significant uncertainties in local corrections for regional/global measurements, as in static corrections for seismic data, shallow refraction and conductor corrections for heat flow and electrical data, etc. For engineering considerations, a high resolution survey with a penetration of about 10 m would be desirable. For geophysical site surveys a lower resolution with greater penetration, of the order a km should suffice. Ideally these surveys should give some three-dimensional resolution of subsurface structure without the need for a dense two-dimensional grid of surface profiles. Technology analogous to side-scan sonar or imaging radar would be useful in this respect.

There are three likely candidates for site-survey geophysical methods, active seismic, electrical or electromagnetic techniques. The magnetic contrast among boulders and regolith, and bedrock/consolidated regolith/regolith without a strong inducing field is likely to be insufficient to produce a useful magnetic signal. Similarly, gravity anomalies associated with long-wavelength, shallow density contrasts or small, shallow density contrasts are unlikely to be resolvable from instrumental noise and deeper source anomalies. Electrical methods could probably be devised with sufficient sensitivity to locate shallow buried boulders, but these techniques would probably require interpretation from an electrical section of the regolith to a structural section. Direct imaging techniques such as reflection seismic or radar techniques would probably have the best resolution for this problem and could be most easily adapted to profile sweeps in a lunar environment.

Both radar and seismic methods have been successful in revealing layering in the lunar regolith. Under the high vacuum conditions of the lunar surface, difficulties may be encountered in coupling a seismic source and receivers to the regolith without stopping the rover and "planting" both source and receivers for each profiling point. Thus, radar probably offers the most attractive option for shallow "rover profiling" with a sideways scanning radar system mounted perpendicular to the rover traverse direction.

Trafficability appears to be primarily restricted by surface topographic features, although problems of raising the loose surface regolith must be considered. These problems appear to be associated with ballistic trajectories associated with surface activities which disturb the surface, and with electrostatic attraction by the lighter particles to charged surfaces.

TERRESTRIAL ANALOGUES

No terrestrial analogue adequately covers the range of properties likely to be experienced in the lunar regolith, but analogues for different physical and mechanical properties may be found. Dry desert conditions in which eolian sands and fluvial gravels are interfingering may present a useful analogue for the lunar regolith for tests of subsurface profiling, both for the detection of boulders and to identify subsurface layering. The properties of the lunar regolith dominated by the fine-sized fraction of the regolith are unlikely to be reasonably represented by any terrestrial regolith materials, primarily due to the different modes of formation of the terrestrial and lunar regoliths. These lunar regolith properties may be simulated by suitably crushed basalt, mixed perhaps with a fine silt component to simulate the effects of the glass component in the lunar regolith.

REFERENCES CITED

- (1) Carrier, W. D., III, Olhoeft, G., and Mendell, W. (1990) Physical properties of the lunar surface. In *Lunar Source Book* (G. Heiken, D. T. Vaniman and B. M. French, eds.), Cambridge University Press, in press.
- (2) McKay, D. S., Heiken, G., Basu, A., Blanford, G., Simon, S., Reedy, R., French, B. M., and Papike, J. (1990) The lunar regolith. In *Lunar Source Book* (G. Heiken, D. T. Vaniman and B. M. French, eds.), Cambridge University Press, in press.
- (3) Alred, J., Bifkin, A., Kennedy, K. J., Petro, A., Roberts, M., Stecklein, J., Sturm, J. (1989) Lunar Outpost. Systems Definition Branch, Advanced Program Office, Johnson Space Center, 60 pp.
- (4) Carrier, W. D., III, Mitchell, J. K., Mahmood, A. (1973) The relative density of lunar soil. In *Proceedings of the Fourth Lunar Science Conference* (Supplement 4, *Geochimica et Cosmochimica Acta*) 3, 2304-2411.
- (5) Taylor, S. R. (1982) *Planetary Science: A Lunar Perspective*, Lunar and Planetary Institute, Houston, 481 pp.

52-33
N91-27105²⁰²¹⁷₁₅

A DESIGN FOR AN INTELLIGENT MONITOR AND CONTROLLER FOR SPACE
STATION ELECTRICAL POWER USING PARALLEL DISTRIBUTED
PROBLEM SOLVING

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

FT 466751

Prepared By:	Robert A. Morris
Academic Rank:	Assistant Professor
University & Department:	Florida Institute of Technology Dept. of Computer Science Melbourne, FL 32901
NASA/JSC Directorate:	Engineering
Division:	Propulsion and Power
Branch:	Power
JSC Colleague:	Mr. Tom Jeffcoat
Date Submitted:	August 21, 1990
Contract Number:	NGT-44-005-803

ABSTRACT

This paper describes work leading to the development of an automated electrical power monitor and controller for spacecraft secondary power systems such as the one anticipated for Space Station Freedom. This work integrates current research in parallel processing and distributed models for problem solving, with current research and development being conducted at NASA-Johnson Space Center on computer automation of spacecraft secondary power distribution and control. Although the emphasis here is on secondary power control, the computational model and representational framework is designed with generality, modularity, and extendibility in mind, and hence could be used in the design of a complete spacecraft monitor and control device.

This research is a response to the need for increased autonomy in spacecraft subsystem control. Electrical Power System (EPS) autonomy is defined to be the capability designed into the EPS which permits onboard task execution through monitoring, computation, command, and control functions with minimal or no ground intervention. The reason for EPS autonomy, especially on based missions with technologically sophisticated spacecraft such as Space Station Freedom, is clear. Already, autonomous load scheduling and fault detection and correction have proved to be plausible applications of intelligent (AI) models for decision making. At the same time, there has been a notable lack of success in integrating intelligent techniques into real time software. To make matters worse, the throughput requirements of electrical power control create even more difficult challenges to automation. It has generally been accepted that for spacecraft applications such as control, advanced software must be integrated with hardware and firmware, as exemplified in recent developments in microprocessor technology. In addition, developments in intelligent systems architecture, such as distributed knowledge representation and parallel processing, should be considered.

The starting point for this research is the Advanced Electrical Power Management Techniques for Space Systems (ADEPTS) project at NASA-JSC. The initial success of ADEPTS has demonstrated the viability of integrating knowledge based techniques for monitoring and control into a parallel processing environment. This research proposes to extend the ADEPTS model with three major goals in mind: first to include more intelligent decision-making capabilities; second, to more fully exploit the computational advantages of parallel processing; and third, to make the ADEPTS model for power management more generic, and hence capable of being applied to future space-bound vehicles such as space station. To fulfill these goals, a cognitive model based on cooperative experts is defined.

INTRODUCTION

The emphasis in this report is on defining a set of communicating processes for intelligent spacecraft secondary power distribution and control. The computer hardware and software implementation platform for this work is that of the ADEPTS project at JSC. The Electrical Power System design which was used as the basis for this research is that of Space Station Freedom, although the functionality of the processes defined here generalize to any permanent manned space power control application.

This report is organized as follows. First, the Space Station Electrical Power Subsystem (EPS) hardware to be monitored is described, followed by a set of scenarios describing typical monitor and control activity. Then the parallel distributed problem solving approach to knowledge engineering is introduced. There follows a two-step presentation of the intelligent software design for secondary power control. The first step decomposes the problem of monitoring and control into three primary functions. Each of the primary functions is described in detail in subsequent sections. This report concludes with suggestions for refinements and embellishments in design specifications.

SPACE STATION SECONDARY ELECTRICAL POWER CONTROL

Monitoring and Control (M&C) involves observing and guiding the behavior of a large physical system towards some objective (ref. 1). Intelligent monitoring and control involves the integration of perceptual, cognitive and reactive knowledge into the monitor and control process. M&C tasks that require such knowledge include: interpretation and prediction of system behavior, reasoned response to important events, diagnosis of exceptional events, and planning long term courses of action.

Secondary electrical power distribution hardware on space station comprises the following elements: DC-to-DC converter units (DDCU), which output 120 volts of DC power, Secondary and Remote Power Distribution Assemblies (SPDA, RPDA) which contain a collection of Remote Power Controller Modules (RPCM) for switching, 1553 buses for transmitting data, power lines, and a set of loads. These elements are arranged in a hierarchical framework to allow for more effective hardware control.

Space station M&C offers challenges which do not arise on either ground-based system control or control on short-term missions such as Space Shuttle. In particular, electrical power will be a limited resource, requiring scheduling on the basis of load priority. In addition, EPS hardware maintenance will need to be performed on potentially dangerous electrical equipment by relatively inexperienced crew members. Finally, the 30-year duration of space station makes continuous ground monitoring

economically and logistically inefficient. Each of these considerations strongly suggest a need for an increased role for automation.

Current design for space station involves dividing the overall EPS M&C process into a hierarchical configuration of tiers (Figure 1). Tier I consists of Operations Management System (OMS) control, including software and hardware for overall ground and onboard management of Space Station operations. Tier II consists of a set of Standard Data Processors (SDP) containing 80386- or 486-based processing capabilities, running concurrently to perform functions such as Guidance, Navigation, and Control (GN&C), data management (DMS), storage unit control, and Node and ITA management. Currently, the Node/ITA SDP and DMS SDP are jointly responsible for Tier II secondary power control. Communication among the Tier II managers and the OMA are via a Fiberoptic Distributed Data Interconnect (FDDI) bus. A Tier III secondary power controller (SPC) provides local control and protection for the various Space Station subsystems. Each SPC, as currently defined, will consist of a Multiplexer-Demultiplexer (MDM), and functions for controlling input/output and self-monitoring. Finally, Tier IV control consists of a hierarchy of Secondary Power Distribution Assemblies (SPDAs) and Remote Power Distribution Assemblies (RPDAs) which are collections of Remote Power Controller Modules (RPCMs). The RPCM switches and distributes 120-Vdc power via a set of unidirectional, solid-state switches. Each RPCM provides deadfacing features, remote manual off capabilities, hardware setpoint programmability and built-in testing features (ref. 2). Each RPCM is assigned a 1553b data bus which responds to external control signals and transmits data and status information within 200 ms of application of control power. These local features of the RPCMs allow for a real time response to fault conditions.

Any proposed advanced automation for space station must integrate with standard control firmware and software. In particular, this report describes software that would naturally migrate to Tier III MDM operations, although certain functions could more properly be assigned to the Tier II level. Certain constraints have been placed on application software at this level, due to two primary factors: the desire to preserve the main function of the MDM, which is high throughput of sensor and effector i/o to the higher control levels, and to maintain compatibility with MDM software. The primary reason for adding intelligence to the MDM is to cut 1553 bus traffic to the higher Tiers (viz., to the SDPs and OMA) (ref. 3). These constraints, among others, are major factors influencing the design decisions made in this report.

To summarize, the intelligent secondary power system controller proposed here, if used on space station, would monitor, evaluate and control performance from the DDCU units to the interface to the loads. It should be viewed as an

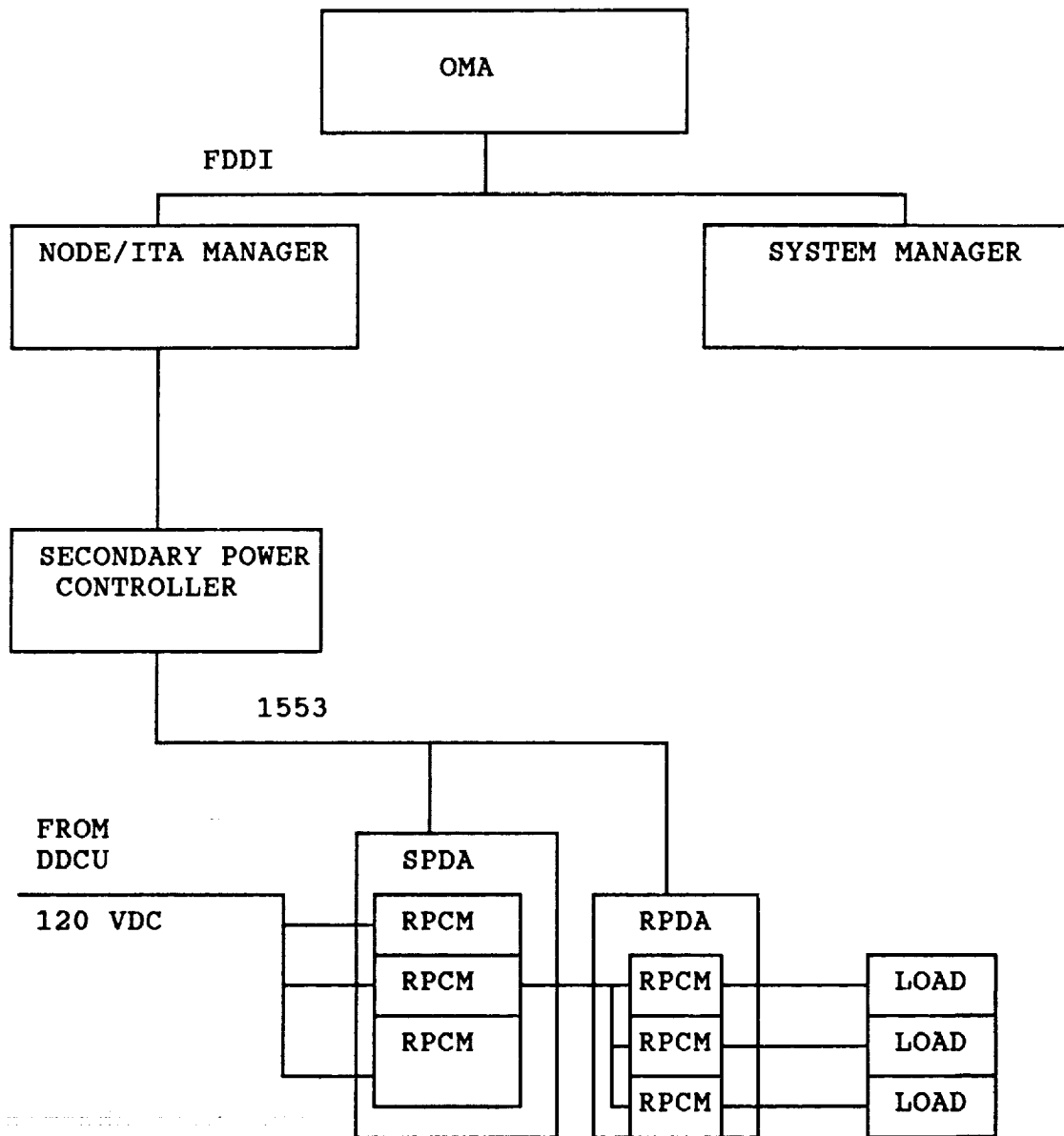


FIGURE 1. SPACE STATION POWER DISTRIBUTION AND CONTROL HIERARCHY

extension of Tier II and Tier III control operations. This controller would therefore have one instantiation for each space station node or element.

SCENARIOS FOR CONTROL

In this section we present four scenarios involving the secondary power distribution hardware for which M&C activities are required. These scenarios are based on a four-fold classification of power system states into normal secure, normal insecure, emergency, and restorative. These states are defined in terms of two sets of constraints: load constraints and operating constraints (ref. 4). Load constraints are requirements for power to the loads. Operating constraints are upper and lower limits on voltage, current, and other parameters of the system.

NORMAL SECURE STATE. In a normal secure state, both operating and load constraints are met. While in this state, load schedules are being dispatched and executed. Status information from the RPCMs and other hardware is monitored and analyzed to check for potential fault conditions.

NORMAL INSECURE STATE. In this state, both load and operating constraints are met by the power system, but there is a likelihood of a transition to an emergency state. Some violations caused by a contingency cannot be corrected without load constraint violations. Preventive rescheduling of loads may be needed to return the system to a secure state.

EMERGENCY STATE. Transition to this state is caused by major violations of operating constraints. Redundant power is currently being used if available, but there is serious loss or degradation of power. While in this state, remedial action or corrective rescheduling is needed to return the system to a normal state.

RESTORATIVE STATE. In this state, only the load constraints are violated, typically because of a need for load shedding and reconfiguration. Actions are required to restore power to the effected loads.

The functionality of the intelligent power controller is based on the need to perform the activities described in these scenarios. It is clear that a number of perceptual, cognitive, and responsive capabilities are involved. When performed by mission control personnel, spacecraft M&C is a cooperative distributed activity performed by a hierarchical committee of flight controllers led by the ``front room'' controllers, with support and backup by power system experts in the ``back rooms.'' In addition, the power system hardware will be

geographically distributed on space station, making distributed, localized on-board automated control a logical approach to intelligent automation. For these reasons, we propose a software model of this activity in terms of a collection of processing agents. This model is presented in the next section.

GENERIC CONTROL MODEL USING DISTRIBUTED PROBLEM SOLVING AND PARALLELISM

Distributed problem solving involves cooperation by a decentralized and loosely connected group of processing units called agents (ref. 5). The problem solving is cooperative in that communication of data and task sharing are required in order to come up with a solution. The problem solving is decentralized in that knowledge, computational resources, control, and data are distributed.

Many distributed problem-solving architectures have appeared in recent years. Among these are blackboard systems, contract net systems, and multi-agent planning systems. These systems are based upon metaphors of social, economic, or biological organization. There are many dimensions along which to categorize these systems. These include:

1. system metaphor (individual, society, committee);
2. grain size (how much processing power is in each of the smallest units);
3. system size (small, medium, large);
4. agent characteristics (controlled or independent, restricted or unrestricted in their resources, interactions simple or complex); and
5. techniques for attaining results (problem solving methods).

Space limitations forbid a detailed discussion of each approach and methodology. Rather, we proceed to present a specific distributed model for intelligent monitoring and control.

For reasons relating to problem domain, implementation, and space station design constraints, the model proposed here is based on the idea of a small, loosely-coupled group of semi-autonomous agents cooperating to maintain the proper distribution of electrical power to each of a set of loads. The actual number of agents will depend on factors such as desired throughput, the number of independent subtasks of the primary tasks that can be discerned, and hardware restrictions. These factors will emerge as the proposed model is refined and instantiated in further research. For now, the proposed model recognizes at least three separate communicating agents.

Partitioning work activity should insure an equal distribution of processing labor; otherwise, the benefits of the distributed approach is lost. The organizational topology of our control model consists of a two-level decomposition of tasks (Figure 2). On the first level, the problem of secondary power

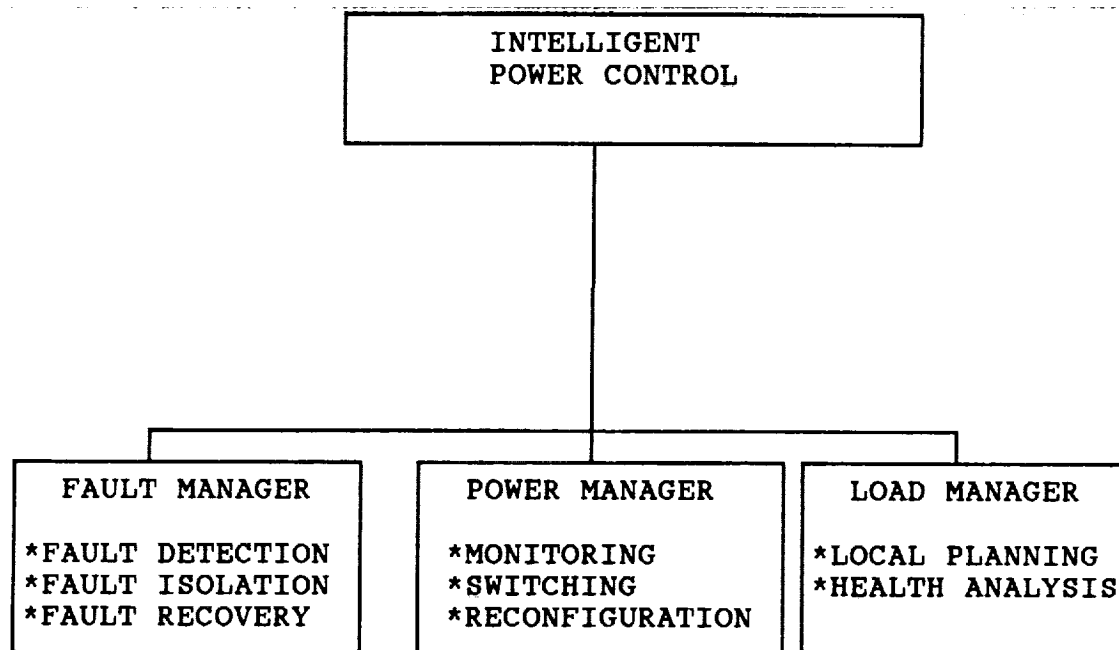


FIGURE 2. POWER CONTROL TASK PARTITIONING

control is divided into three primary tasks: power management, fault management, and load management. This division is based on a decomposition of monitoring and control functions in terms of a three-fold temporal partitioning of the events that make up the behavior of the power system. Events that occurred in the past include exceptional or urgent events that require diagnosis, and are of interest to the fault manager. Events that are presently occurring include the flow of electrical power and relay switching. These events need to be monitored and controlled, and will be the assigned task of the power manager. Finally, future events are those which need to be planned or predicted; in particular, loads need to be scheduled and possible unhealthy conditions predicted. This will be the job of the load manager.

On the second level, each primary task is subdivided into a collection of perceptual and cognitive operations. For example, the fault manager consists of procedures for gathering and evaluating voltage and current information in order to perform diagnostics. Each of these subfunctions are described further in the next section.

As mentioned, in this report, we will view each primary task as an intelligent agent. Further refinements of this model may result in assigning to each primary task a network of agents, depending on whether certain subtasks within a primary task may be viewed as independent perceptual or cognitive activities. Furthermore, each primary agent can be viewed as either operating concurrently on a separate processor, or as sharing processing resources. In the discussion that follows, we will assume for simplicity that the agents run continuously on their own processors. This assumption allows for a consideration of the potential for parallelism in the monitor and control process.

Two sorts of parallelism can be distinguished: in one, a set of necessary and independent tasks are identified and processed concurrently (AND-parallelism). The other (OR-parallelism), is the result of the non-determinism involved in intelligent problem solving, i.e., in the potential for solving the same problem in a number of different ways. As will be shown, both kinds of parallelism are inherent to the problem of monitoring and control.

As mentioned, in choosing this preliminary model, a number of constraints have been applied. Some of these involve integrating distributed conventional control hardware and software with intelligence. Other constraints involve the nature of the process of spacecraft monitoring and control itself, which, when solved by humans, typically involves a small group working cooperatively on the problem. Finally, the desire to implement the model using the transputer environment, as envisioned by the ADEPTS designers, was a consideration in constructing the model.

PARTITIONING KNOWLEDGE AND EXPERTISE AMONG LOCAL AGENTS

This section contains a generic description of each of the three primary agents in terms of four characteristics: the agent's functional components; responsibilities; resource requirements (software, hardware, and communication); and interactions (i.e., which agents talk to which others, and what do they exchange). In addition, there is a brief informal discussion of the potential for AND- and OR-parallelism within each agent's activity. Finally, we describe briefly how each of these generic functions could be instantiated on space station.

Local Fault Management

Responsibilities. A fault manager is responsible for isolating, identifying, and aiding in the recovery from faults involving power system components. The problem of fault diagnosis can be stated thusly: given a set of symptoms, find a set of faults to account for them. The problem of fault identification is complicated by the fact that different faults may share symptoms, and also by virtue of the possibility of multiple faults. The amount of reasoning skills required by fault diagnostic software is inversely proportional to the amount of sensor data to be processed. Therefore, by adding an intelligent fault manager to system software, data processing can be potentially reduced.

Functional Components. Every intelligent fault manager consists of the following:

1. A procedure for gathering data
2. A simulator of the behavior of the artifact
3. A diagnostic problem solver, consisting of:
 - i. A procedure for recognizing discrepancies
 - ii. A procedure for generating hypotheses
 - iii. A procedure for testing hypotheses

To perform fault management, data must be constantly gathered and tested against the expectations generated by the simulator. When a discrepancy is identified, a symptom set is generated. The model is consulted to generate hypotheses about the cause of the discrepancy. Each hypothesis is tested by observing further data, or by injecting the hypothesis into the simulated model of the system, generating simulated symptoms which are further matched to the actual data. This procedure is repeated until a unique hypothesis is generated.

Resource Requirements. Input to this agent will consist of voltage and current data from remote controller modules. The knowledge sources of this agent include the model of the components of the subsystem under its control, including the buses, switchgear, and loads. Its primary local data are current symptom sets.

Interactions. This agent can perform its major functions locally. Diagnostic information, i.e., data identifying faulty components; the nature of the fault; and recommendations for recovery procedures; should be communicated to the local power manager, as well as the load manager. The load manager must use this information in plan generation. In addition, the human user of the system should be alerted and recommendations for recovery enumerated.

Potential for Parallelism. AND-parallelism is inherent in the relative independence of the subtasks that are performed by the fault manager. For example, the data analysis and problem solving phases can be run concurrently. The fault manager activity also allows for some degree of OR-parallelism, e.g., when multiple hypotheses are generated from symptom sets. Each hypothesis can be tested separately against current data. This parallelism could be useful if multiple fault detection is required by the fault manager.

Space Station Instantiation. Distributed diagnosis is particularly suited to space station design. The point of local diagnosis is to reduce the flow of status data from the component controllers to the upper level managers. Furthermore, the purpose of intelligent automated diagnosis is to reduce the number of points that need to be sensed in order to make a diagnosis, thus reducing the data processing requirements.

Power Manager

Responsibilities. The power manager is responsible for maintaining the proper flow of electrical power to the loads. This involves controlling the allocation of power resources to meet the current system load. The cognitive requirements for the power manager consist in the ability to compute changes in configuration of the system on the basis of load priority, which includes the ability to determine which loads must be shed. It is also envisioned that the power manager will share the local monitoring tasks with the fault manager.

Functional Components. The activities of the power manager can be broken down into the following functions:

1. Configuration Processor
2. Resource Controller
3. Health Monitor

The configuration processor computes switch settings for the remote power controllers based on current load demands and load priorities. These load demands may either result from normal execution of a load schedule, or by an exceptional event such as a component fault or addition of a load. For example, if power to a critical load has been switched automatically to a redundant path, and if a part of that path has already been fully allocated to other loads, then the power manager will compute which loads to shed along the redundant path.

The other functions of the power manager may consist of more conventional algorithmic software. The resource controller distributes commands to remote controllers (RPCMs) for switch settings. The health monitor will periodically perform analysis of the data from the remote power controllers.

Resource Requirements. Like the fault manager, the power manager will contain component and behavior knowledge of the power system. In addition, it will require at least two small databases: one, consisting of current switchgear configuration information, which will be updated as load requirements change; the other, a temporal database, which will allow for the maintenance of health information. Finally, it must contain load priority data, in order to compute power allocation.

Interactions. The power manager will interact both with the fault manager, as described above, and the load manager. The load manager will request information about current power availability in order to verify or update a load schedule. In addition, configuration updates will be reported to users.

Potential For Parallelism. The three power manager subtasks are relatively independent, which suggests a source of AND-parallelism. In particular, the monitoring task can be performed concurrently with the other tasks. OR-parallelism may be found in the computation of configurations.

Space Station Instantiation. The power manager will allow for more load shedding decisions and health monitoring decisions to be made locally, thus reducing the volume of OMA-directed activity.

Load Manager.

Responsibilities. Scheduling loads can be a highly complex problem. Distributing this responsibility by localizing the processing would lessen the bottleneck associated with generating load schedules. A promising approach to partitioning the problem solving knowledge and control is by distinguishing between the static and dynamic aspects of scheduling. Static scheduling involves taking a set of loads and constraints and generating a consistent, usable schedule. Parameters and constraints which are used to develop a static schedule involve global knowledge of each EPS function, from generation and storage to distribution. As such, the design and implementation of the static scheduling procedure would occur on Tier I or Tier II level control. Consequently, this problem is beyond the current scope of the subsystem controller being designed here.

The dynamic scheduling procedure has the function of modification of proposed schedules on the basis of the addition or deletion of resource or load constraints. These new constraints may arise as the result of damage to EPS equipment, additional load requirements, or other resource constraints such as crew or instrument availability. This, we propose, is a function that can be localized and distributed to a Tier III-

type controller. This dynamic load scheduler will be called a (local) load manager.

Functional Components. A load manager will verify and modify schedules and plans based on current resources. The need for such a function arises due to expected frequent changes in resource availability. Schedule verification consists of ensuring that the power and other resource expectations demanded by the static schedule can be currently met. For example, the static schedule may assume that a certain activity A, can be assigned an interval of time T with power requirements P, instrument I and crew members C D and E. The contingency handler will verify that no changes to any of these resource requirements has occurred since the static schedule was generated. If current resources do not match load requests, the load manager will attempt to modify the schedule in order to satisfy the requests.

Resource Requirements. The data and knowledge sources for this module include a load profile database for each load, a resource database of currently scheduled resources, and a database consisting of current schedule information.

Interactions. Rescheduling may involve reconfiguration and load shedding. Therefore, it will be necessary for the load manager to request data and share tasks with the power manager. It will also receive and send schedule information from/to human planners.

Potential for Parallelism. Scheduling can be viewed as a search problem. Parallel versions of depth-first search algorithms are being developed, which are essentially examples of exploiting OR-parallelism.

Space Station Instantiation. Because of the complexity of space station planning, it is less clear to what extent a local planner can be employed on Tier III-level software. It is likely that automated planning on station would be a Tier I or II activity. It was suggested that a Tier III load manager could be an "intelligent interface" to human planners, notifying them that planned loads can or cannot be powered, and perhaps suggesting contingency short-term schedules.

Summary of Preliminary Model

The model described here is in a preliminary stage of development. More is required in order to define the knowledge sources of each agent, the communication and synchronization mechanism among the primary agents, and the potential for concurrency among the agents' subfunctions.

A useful metaphor for distributed power control is human organization. To avoid control bottlenecks among upper level managers, it is useful to distribute intelligent decision making among lower-level managers. On the other hand, too much local control could result in the lack of a common direction among the different local units. The issue of the amount of local

automated control will need to be addressed in applying the generic model proposed here to space station.

SUMMARY AND FUTURE DIRECTIONS

This paper has sketched a design for an intelligent secondary power controller, based on current Space Station Freedom secondary power distribution hardware specifications, as well as on current developments in cooperative problem solving methodology. The model has arisen from conversations with station designers of EPS hardware, the OMS, data management system (DMS), and application software. These developers were helpful in detailing the functionality of the different software modules, and suggesting avenues for intelligent automation.

A number of improvements and extensions of this work are needed. First, as mentioned, the model is in need of significant further development. A prototype of the model should be built and tested, and ported to the ADEPTS transputer implementation environment. Currently, an implementation of a prototype of this model is being developed.

Second, more agents should be added as needed. For example, a separate data pre-processing agent could monitor sensors and translate and filter sensed data to the appropriate manager. In addition, an intelligent user interface could be added for displaying data to users, such as crew members, with varying expertise in the problem domain.

Third, more functional autonomy could be added to each agent. Each agent could possess more sophisticated cognitive and decision-making capabilities. Our design favored a more modest approach to agent autonomy, in line with current opinions regarding SPC MDM processing capabilities. As early prototypes are tested and validated, more functionality can be added.

Fourth, a more detailed investigation of the potential for AND- and OR-parallelism should be undertaken. Finally, the committee of agents defined in this report could be extended to include more global (i.e., Tier I or Tier II) control operations, such as static scheduling and user interfacing. With few modifications, the generic model could also be instantiated for primary as well as secondary control operations.

It is believed that the model defined here could be useful to space station designers in the automation of space station power control. Instantiating the generic model for space station will imply tailoring the roles of each manager to the requirements of a Tier III or Tier II controller. As the model proposed here and the space station automation process progress, the proposed software could 'migrate' on board in later growth stages. Finally, the overall model advanced here can be applied to future automation projects such as for lunar and Mars bases.

ACRONYMS

ADEPTS	Advanced Electrical Power Management Techniques for Space Systems
DDCU	DC-to-DC Converter Units
DMS	Data Management System
FDDI	Fiberoptic Distributed Data Interconnect
FDIR	Fault Detection, Isolation and Recovery
ITA	Integrated Truss Assembly
M&C	Monitoring and Control
MDM	Multiplexer/Demultiplexer
OMA	Operations Management Applications
RPCM	Remote Power Controller Module
RPDA	Remote Power Distribution Assembly
SDP	Standard Data Processor
SPC	Secondary Power Controller
SPDA	Secondary Power Distribution Assembly

REFERENCES

1. Hayes-Roth, B.; Washington, R.; Hewett, R.; Hewett, M.; and Seiver, A.: Intelligent Monitoring and Control, (Proceedings of the International Conference on Artificial Intelligence, IJCAI-89), Morgan Kaufmann Publishers, Inc., 1989, pp. 243-249.
2. Interface Development Document: Remote Power Controller Module Standard (Preliminary), Space Station Freedom Electrical Power System, Work Package 04 Document, SSP 44004, March 20, 1990.
3. Systems Engineering and Integration Data: Multiplexer/Demultiplexer (MDM) DMS PDR#3 Support Data, Work Package 2 (WP-2) MDC H4778, March, 1990.
4. Talukdar, S.; Cardozo, E.; Leao, L.: Toast: The Power System Operator's Assistant, Proceedings of the IEEE, Vol. , No. July, 1986, pp. 53-59.
5. Gasser, L.; Huhns, M.H.: Distributed Artificial Intelligence Volume II, Morgan Kaufmann Pub., Inc., 1989.

S3 -37
N91-27106-8
P-13

**FUZZY LOGIC CONTROL SYSTEM TO PROVIDE AUTONOMOUS
COLLISION AVOIDANCE FOR MARS ROVER VEHICLE**

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

H 2086788

Prepared By:	Michael G. Murphy, Ph.D.
Academic Rank:	Professor
University & Department:	University of Houston - Downtown Department of Applied Math. Sciences Houston, Texas 77002

NASA/JSC

Directorate:	Information Systems
Division:	Information Technology
Branch:	Software Technology
JSC Colleague:	Robert N. Lea, Ph.D.
Date Submitted:	August 10, 1990
Contract Number:	NGT-44-005-803

ABSTRACT

NASA is currently involved with planning unmanned missions to Mars to investigate the terrain and process soil samples in advance of a manned mission. A key issue involved in unmanned surface exploration on Mars is that of supporting autonomous maneuvering since radio communication involves lengthy delays. It is anticipated that specific target locations will be designated for sample gathering. In maneuvering autonomously from a starting position to a target position, the rover will need to avoid a variety of obstacles such as boulders or troughs that may block the shortest path to the target. The physical integrity of the rover needs to be maintained while minimizing the time and distance required to attain the target position.

Fuzzy logic lends itself well to building reliable control systems that function in the presence of uncertainty or ambiguity. Systems already developed by Robert N. Lea of NASA/Johnson Space Center and others test out well for shuttle proximity operations and for Mars rover trajectory control in the absence of obstacles. One of the tools investigated for developing a fuzzy logic obstacle avoidance control system is the Togai InfraLogic TILShell fuzzy logic expert system shell that provides a software development tool to express fuzzy logic rules as well as crisp rules and procedural code segments, then generates standard C code for implementation. This shell has been successfully used in the proximity operations and rover trajectory systems. Fuzzy logic control systems have been underutilized in American application areas but have been very successfully employed in a variety of commercial and industrial applications in Japan. An additional technology that shows promise of applicability to a variety of control system environments (such as the one addressed here) is that of learning and adaptation based on neuronlike elements.

This report addresses three major issues:

1. the nature of fuzzy logic control systems and software tools to implement them;
2. collision avoidance in the presence of fuzzy parameters; and,
3. techniques for adaptation in fuzzy logic control systems.

In addition to current work on these issues, possible directions for future work are given.

INTRODUCTION

The Space Exploration Initiative of the United States includes plans for unmanned missions by NASA to Mars to investigate the terrain and to process soil samples in advance of a manned mission. Path planning is a crucial element in providing for this exploration. More specifically, surface exploration of Mars involves an environment where uncharted obstacles are likely to interfere with maneuvering of a rover vehicle and where radio transmissions to and from Earth involve lengthy delays, hence the need for autonomy of operation of the vehicle. In moving autonomously from a starting position to a target position, the rover will need to avoid a variety of obstacles such as boulders or troughs that may block the shortest path to the target. The problem is to maintain the physical integrity of the rover while minimizing the time and distance required to attain the target position.

Although the problem of autonomous path planning including collision avoidance is not new, the research into the use of fuzzy logic in the decision and control process is new. The theory of fuzzy sets and fuzzy logic has matured since 1965 when it was introduced by Professor L. A. Zadeh of the University of California, Berkeley (1). The book by Klir and Folger (2) gives a good treatment of the fundamentals of this field. What has been lacking until the last few years are significant applications of this body of knowledge to practical endeavors in business and industry. This lack has been partially addressed by Japan, where a variety of applications involving fuzzy logic have been developed. One of the most important types of application has been fuzzy logic control systems (3,4). Togai Infralogic, Inc., of California has recently marketed a fuzzy logic expert system shell to facilitate the production of software for fuzzy logic applications(5). Robert N. Lea of NASA/JSC and others have been working on fuzzy logic applications to space vehicle rendezvous and proximity operations (6,7,8). Since autonomous path planning with collision avoidance is structurally a control problem and since fuzzy parameters are likely to be present in the Mars rover maneuvering environment, a fuzzy logic control system would be a promising avenue of investigation.

FUZZY LOGIC CONTROL SYSTEMS AND SOFTWARE TOOLS

The purpose of control systems is to achieve and then maintain a desired state of a process by monitoring the output and state of the process then taking responsive control actions. An example would be monitoring the temperature and pressure of a chemical process then raising, maintaining, or lowering the temperature to achieve or keep an optimal state for the process. Since temperature and pressure are likely to change during operation, this monitoring needs to be continued at regular time intervals. It is also desirable that the control be automated so that operator intervention is minimized (but permitted).

Traditional automatic controllers usually require a sufficiently precise mathematical model of the process. For many complex control environments, it is very difficult if not impossible to construct a satisfactory model due to issues such as nonlinearity, time variance, or poor quality of available measurements.

In cases such as these, a human operator may be able to effect superior control over traditional automatic controllers. An expert human operator consciously or unconsciously uses heuristic control rules or "rules of thumb" based on experience to control the process. When articulated by the operator, these rules usually involve linguistic descriptions with vague or ambiguous concepts. An example could be "if the speed of a conveyor belt is very fast, then the likelihood of an accident is high" with the terms "very", "fast", and "high" being defined only approximately. Terms such as "fast" in this context are considered values of the linguistic variable "speed", where the degree of membership μ could be given by means of a graph, such as the one in Figure 1 below, where the horizontal axis takes on possible values for speed, and the vertical axis gives the degree (0 to 1) that a particular speed is a member of the linguistic value above it. Of course, the membership function could be of any form and not necessarily piecewise linear. The term "very" would usually be interpreted as the square of the membership of "fast" in our context.

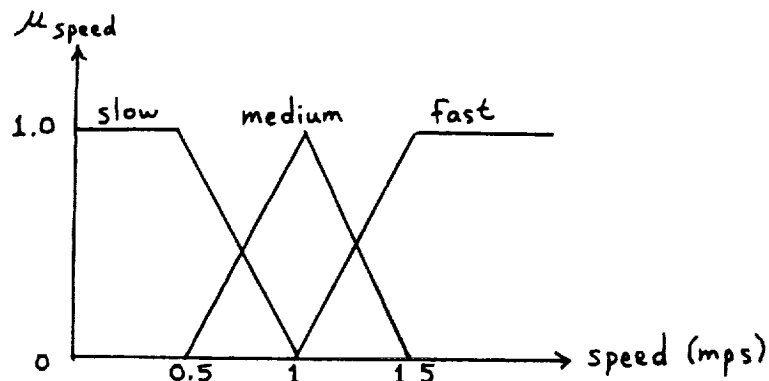


Figure 1. - Graphical representation of linguistic values of "speed" for conveyor.

A rule-based fuzzy logic controller consists of a finite set of fuzzy control rules which are processed by applying what is called approximate reasoning. A typical fuzzy logic control rule is of the form:

If x is A and y is B, then z is C.

Figure 2 below indicates how the memberships of x in A and y in B yields the membership level of z in C.

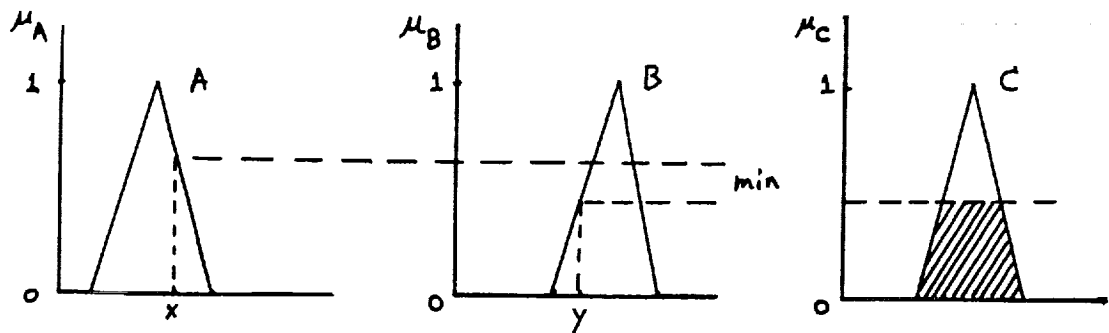


Figure 2. - Graphical representation of approximate reasoning for one rule.

One of the ways used to combine the recommended control actions (in membership form) for several rules is to determine the center of area for all of the actions, thus giving a crisp control action as indicated in Figure 3 below. This is sometimes called conflict resolution or inferencing.

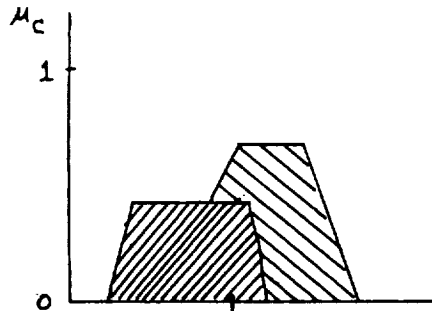


Figure 3. - Conflict resolution (inferencing) for more than one control rule.

Figure 4 below gives the architecture for a typical fuzzy logic control system. Simple systems of this type are usually called fuzzy logic controllers.

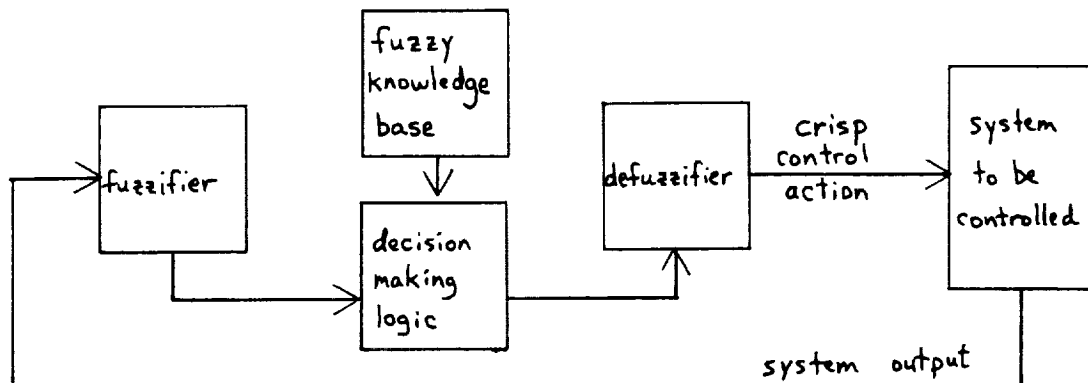


Figure 4. - Basic architecture for fuzzy logic control.

In addition to straightforward programming from scratch and analog fuzzy logic chips, there are now a limited number of software tools to facilitate the development of fuzzy logic expert systems, with a fuzzy logic control system as an example of such. One of the most readily available and most conveniently used tools is the fuzzy logic expert system shell called TILShell from Togai InfraLogic Inc. (5). Technically, TILShell is a graphical editor that creates a database for Togai InfraLogic Inc.'s Fuzzy-C Compiler, which then generates C language code with the data structures and inference engine required to implement the system. This code may then be compiled and linked to other application code. Additional software/hardware hybrid tools are being developed by Apt Instruments Inc. and should be available by the early part of 1991.

Fuzzy logic control systems have been implemented effectively for a variety of applications in Japan. A video teleconference on July 24, 1990, linked the close of the International Conference on Fuzzy Logic and Neural Networks (IIZUKA'90 at the Kyusu Institute of Technology in Japan) with interested parties at the Johnson Space Center, and it was made clear that this is very likely the dawn of an era of vigorous and expanded growth in fuzzy logic applications.

COLLISION AVOIDANCE UNDER CONDITIONS OF UNCERTAINTY

This section represents an original approach to collision avoidance under conditions of uncertainty. The first case to be considered for an avoidance system is a fuzzy obstacle that is not moving. The objective is to have the vehicle maneuver from a source position to a destination position (for sample gathering, etc.) in the shortest amount of time and avoid collision with the obstacle. Figure 5 below gives the scenario for this problem.

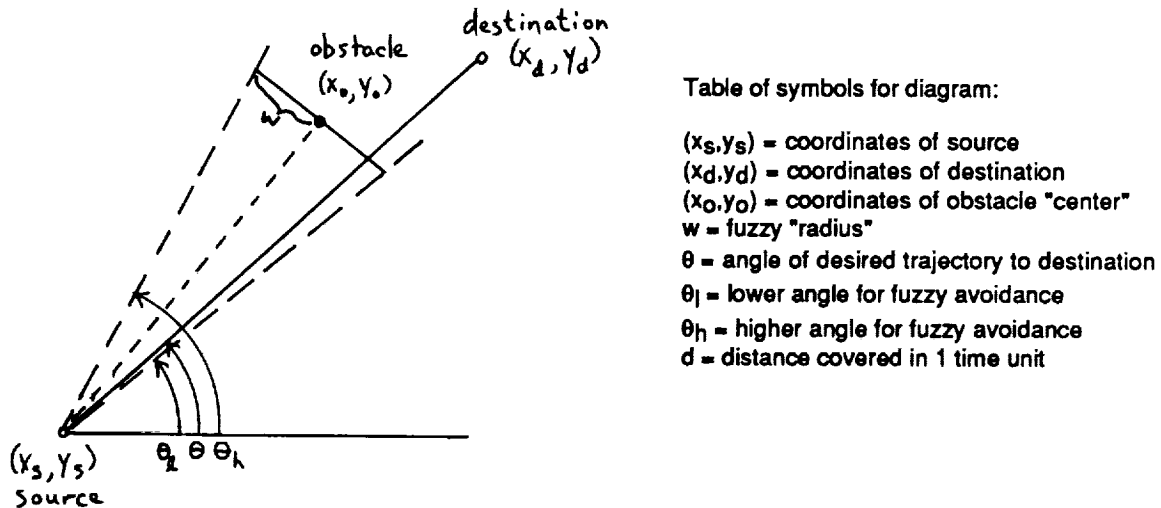


Figure 5. - Avoidance of a static fuzzy obstacle.

The symbol "w" is used to denote the fuzzy "radius" of an obstacle; that is, the radius about the obstacle "center" for which the membership in "is an obstacle" is above some threshold (e.g., 0.1). Both the fuzzy "radius" and the fuzzy "center" for the obstacle are assumed to be available from current sensor data. This represents a significant relaxation of requirements for object identification and object extent determination. In general, the objective is to take fuzzy information regarding a potential obstacle, determine whether it is an obstacle, take evasive action if necessary, then continue until the destination is reached. The algorithm for collision avoidance in the case of this static fuzzy obstacle is:

1. Receive (x_0, y_0, w) from sensor processing.
2. Check for obstruction:
 - if $\theta_l < \theta < \theta_h$
 - then change θ to θ_l if $|\theta - \theta_l| < |\theta - \theta_h|$,
 - otherwise change θ to θ_h
 - else no change in θ .
3. Advance vehicle one time unit :
 - new $x_s = \text{old } x_s + d \cos \theta$
 - new $y_s = \text{old } y_s + d \sin \theta$.
4. Repeat steps 1., 2., 3. until destination is reached.

Once the vehicle is sufficiently close to the destination, it may be necessary to enter a new control environment for positioning, etc., so step 4. above may actually cease when some critical minimum distance to the destination is reached.

The second case to be considered for an avoidance system is an amorphous fuzzy obstacle system that is not moving. The term "amorphous" in this setting means that there may be more than one obstacle or perhaps even a mass of obstacles that have not been separated or may not be separable. Once again this represents a simplification of the requirements for sensor processing while providing a satisfactory environment for collision avoidance. The objective again is to have the vehicle maneuver from a source position to a destination position (for sample gathering, etc.) in the shortest amount of time and avoid collision or damage to the vehicle. Figure 6 below gives the scenario for this problem.

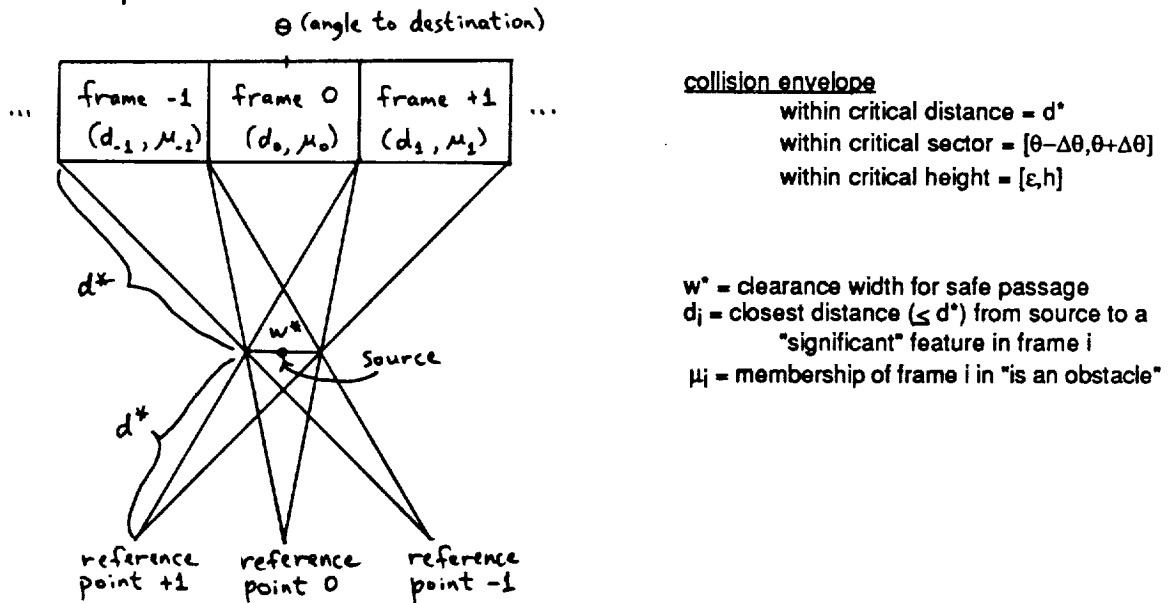


Figure 6. - Avoidance of a static amorphous fuzzy obstacle system.

The idea in this avoidance process is to provide a corridor for safe passage of the vehicle. If θ is the angle from the source to the destination, then the angular interval $[\theta - \Delta\theta, \theta + \Delta\theta]$ is broken up into an odd number of angular sectors of width at least $2w^*$. The 0th sector is centered about θ at a distance of d^* from the source on the way to the destination. The first sector to the right of center is labelled 1, then the second sector to the right is labelled 2, and so forth. The first sector to the left of center is labelled -1, then the second sector to the left is labelled -2, and so forth. The sensor data is processed for each sector by scanning the region from the source to the sector boundary as indicated in Figure 6. The reference point is moved back from the source position to a position d^* units behind the source along a line extended from the center of the sector through the center of the vehicle. This ensures that the entire sector path is evaluated for safe passage. Each sector may be considered a frame of reference which is scanned with sensors and assigned a distance and membership in "is an obstacle" which is a generalized form of obstacle detection. In essence, what happens is that the closest distance to a "significant" obstacle is determined; significant would mean above a certain threshold of membership. The exact form of sensor processing is not addressed here. A simplified version of the algorithm for collision avoidance in the case of a static amorphous obstacle system is:

1. Receive (d_i, μ_i) for each frame from sensor processing.
2. Select path if possible:
 - if $\mu_0 < \text{threshold}$
 - then proceed in direction θ
 - else pick i with smallest $|i|$ and $\mu_i < \text{threshold}$
and proceed in direction of center of frame i
 - else expand angular interval by $\Delta\theta$ on both
sides and check for safe corridor there
until one is found.
3. Repeat steps 1., 2. until destination is reached.

Once the distance to the destination is sufficiently close (e.g., $< d^*$), it is likely that a proximity positioning subsystem would take over for the final approach. One rather natural adjustment to the algorithm would be to scale the membership value of each frame j with a "significant" obstacle by d_i/d_j where i is the frame with the closest "significant" obstacle. Another feature to consider adding is that of speed control (e.g., speed up for safer passage such as the 0th sector with a low membership value and slow down for riskier passage such as a higher numbered sector with a membership value near the threshold). A simplified form of this obstacle system and algorithm has been implemented by Larry Walters of Lincom Corporation using the TILShell software tool. Further work is needed to make a verifiable autonomous avoidance system for the case of a static amorphous fuzzy obstacle system.

The third case considered is a single fuzzy obstacle with fuzzy movement. Work to date on this case is limited and represents a preliminary framework upon which to build in the future. The situation is that an obstacle with an estimated "center" and fuzzy "width", fuzzy magnitude of speed, and fuzzy direction may cross the path of the vehicle. The problem is to generate the range of positions that the obstacle may take and then determine if the vehicle is likely to be endangered; if so, evasive action must be taken. Figure 7 gives the scenario for the problem. This is a much more complicated problem than that of a single fuzzy obstacle, and it may be that the amorphous static approach may provide the basis of an approach for this situation as well.

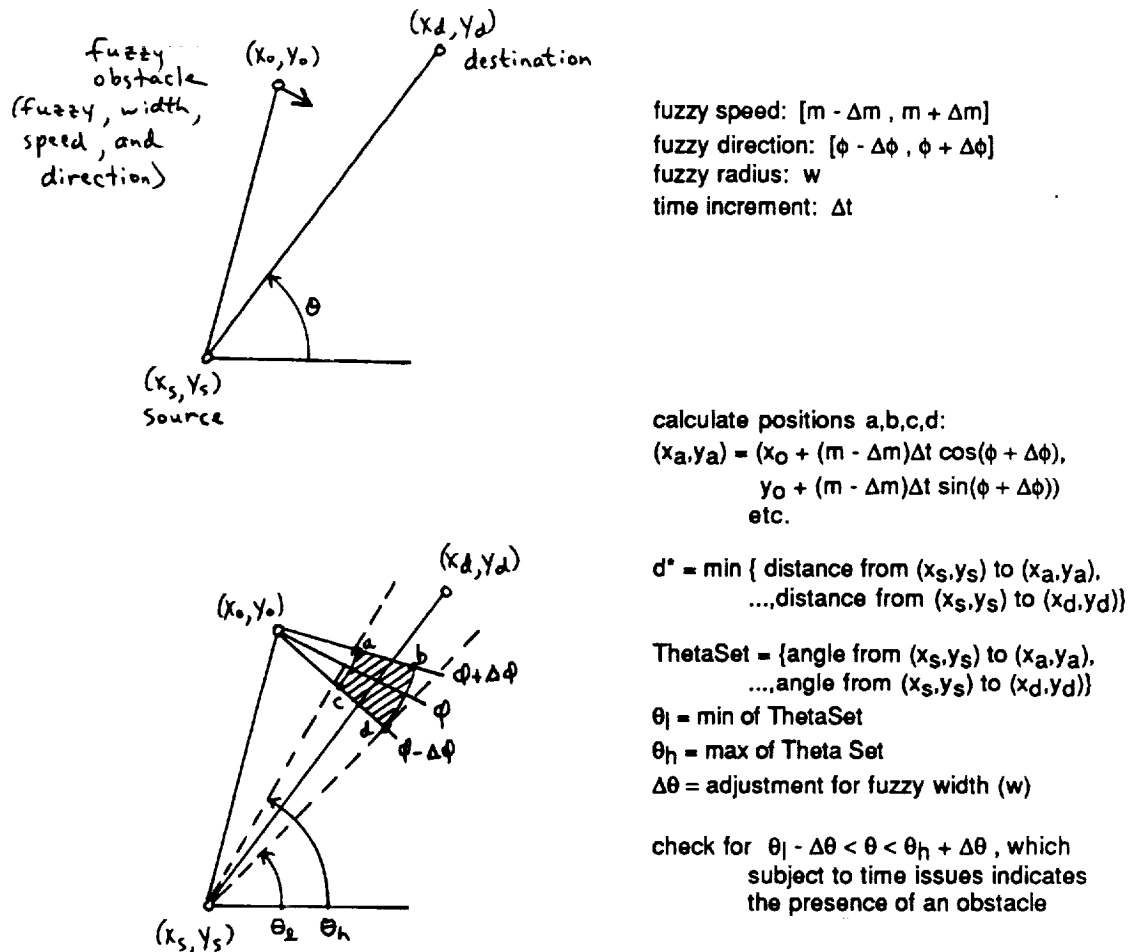


Figure 7. - The situation for avoidance of a moving fuzzy obstacle.

ADAPTATION IN FUZZY LOGIC CONTROL SYSTEMS

The issue of developing adaptive fuzzy logic control systems is important and promises to permit the building of control systems that operate under conditions of uncertainty or ambiguity while projecting robustness of operation as conditions change and the likelihood of graceful degradation of operation under loss of control functionality. Adaptive in this setting means developing the system adaptively from training sets of typical data or even permitting "on the fly" modification of the system to reflect changing environments as they occur. The book by Widrow and Stearns (9) provides a framework for adaptive systems in general while the book by Kosko (10) addresses environments closer to those in this report including fuzzy logic systems and adaptive techniques associated with neural networks. Adaptive rule-based control systems that use neuronlike elements in a crisp environment have been addressed by Barto, Sutton, and Anderson (11) and then extended to systems with fuzzy logic rules by Lee and Berenji (12, 13, 14, 15).

The architecture for one type of trainable adaptive fuzzy logic control system is given in Figure 8. It makes use of coupled neurons that learn based on system performance, thus updating the rule-base for enhanced performance. The flow of calculations is given in Figure 9. A good reference for this approach is the paper by Lee (13).

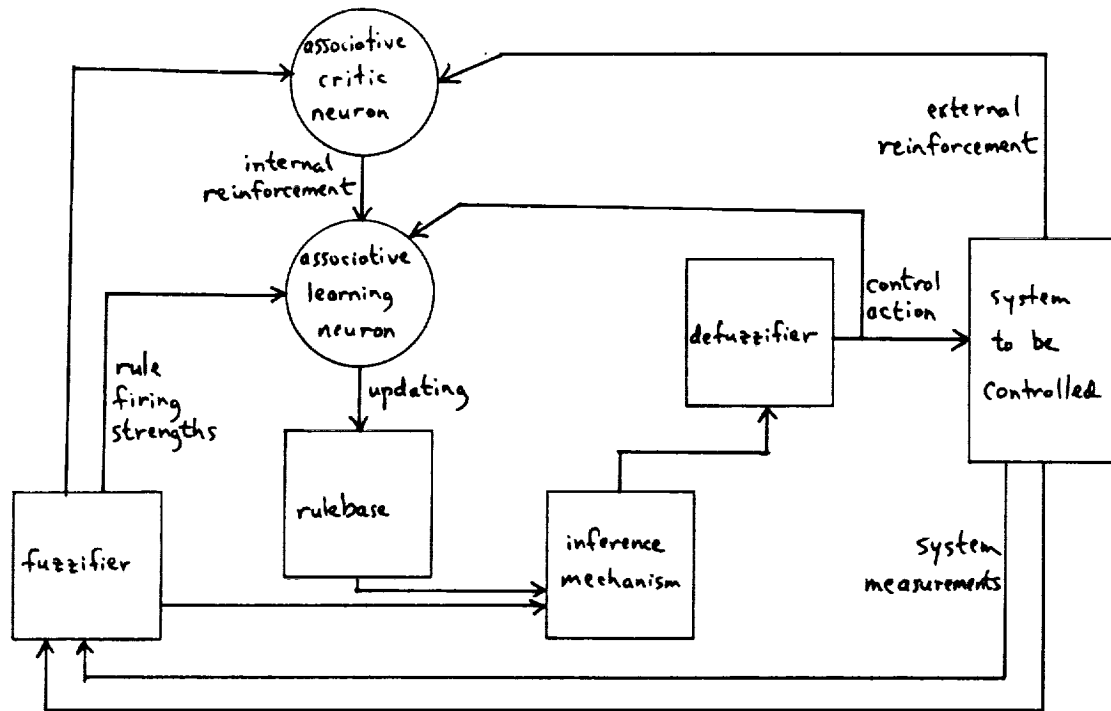


Figure 8. - Architecture for adaptive fuzzy logic control system.

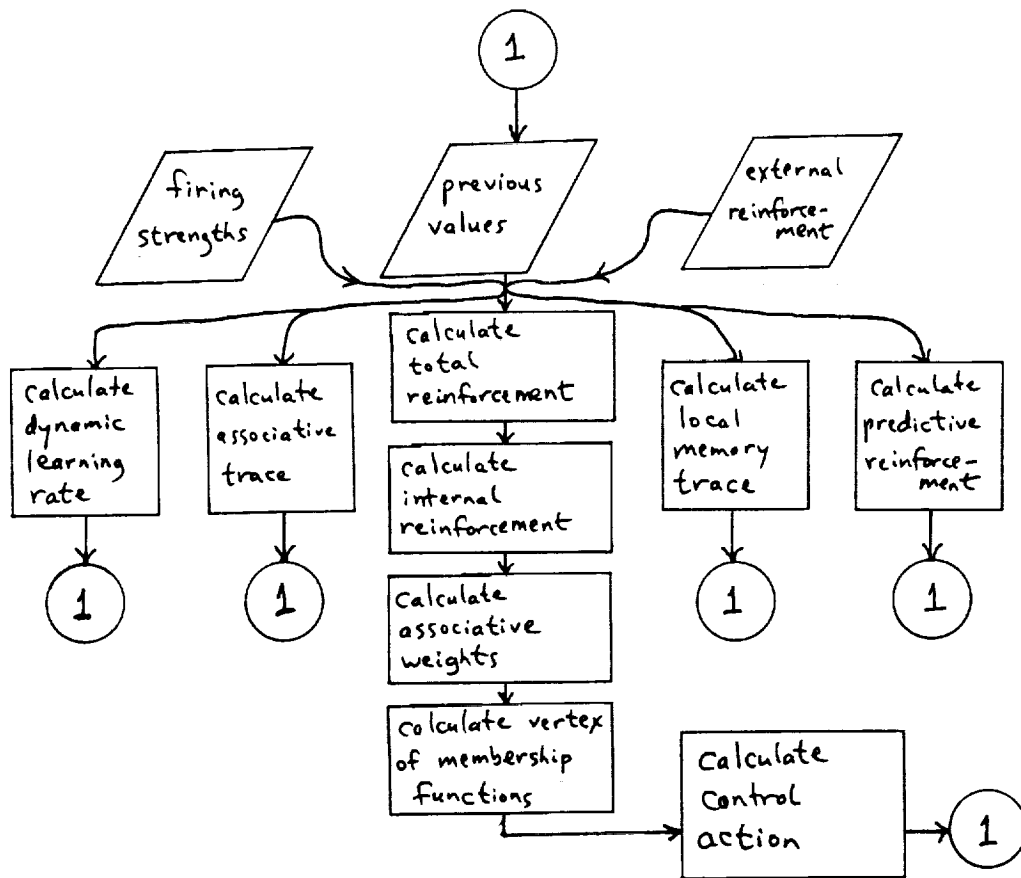


Figure 9. - Flow of calculations for adaptive fuzzy logic control system.

CONCLUSION AND FUTURE DIRECTIONS

Based on the literature identified this summer and additional topics explored from scratch, it is clear that the area of fuzzy logic control is one of significant opportunity for applications. In particular, it is likely to work well under conditions of vagueness or ambiguity and in settings where autonomy is important since fuzzy logic control supports approximate reasoning and model simplification. The specific area of collision avoidance for a Mars rover is a natural setting for the application of these techniques. Further work is needed to explore fuzzy logic collision avoidance in more general situations than those explored in this report. Examples would be fuzzy obstacles with fuzzy movement, collision avoidance in three-space (e.g., involving the space shuttle or other non-surface space vehicles), and developing a working adaptive mechanism for collision avoidance.

REFERENCES

1. Zadeh, L. A.: Fuzzy sets. *Information and Control*, vol. 8, pp. 338-353.
2. Klir, George J.; and Folger, Tina A.: Fuzzy Sets, Uncertainty, and Information. Prentice-Hall, 1988.
3. Lee, Chuen Chien: Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part I. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 2, March/April, 1990, pp. 404-418.
4. Lee, Chuen Chien: Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part II. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 2, 1990, pp. 419-435.
5. TILShell User's Manual, Release 1.1. Togai InfraLogic Inc., Jan., 1990.
6. Lea, Robert N.: Automated Space Vehicle Control for Rendezvous Proximity Operations. *Telematics and Informatics*, vol. 5, no. 3, 1988, pp. 179-185.
7. Lea, Robert N.: Applications of Fuzzy Sets to Data Monitoring and Process Control. *Proceedings of the ISA'88 International Conference and Exhibit*, October, 1988, pp. 1495-1502.
8. Lea, Robert N.; Togai, Masaki; Teichrow, Jon; and Jani, Yashvant: Fuzzy Logic Approach to Combined Translational and Rotational Control of a Spacecraft in Proximity of the Space Station. *Proceedings of the Third International Fuzzy Systems Association Congress*, August, 1989.
9. Widrow, Bernard; and Stearns, Samuel D.: Adaptive Signal Processing. Prentice-Hall, 1985.
10. Kosko, Bart: Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Prentice-Hall, 1990.
11. Barto, Andrew G; Sutton, Richard S.; and Anderson, Charles W.: Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, no. 5, 1983, pp. 834-846.
12. Lee, Chuen-Chien; and Berenji, Hamid R.: An Intelligent Controller based on Approximate Reasoning and Reinforcement Learning. *Proceedings of the IEEE International Symposium on Intelligent Control 1989*, pp. 200-205.

13. Lee, Chuen-Chien: Intelligent Control Based on Fuzzy Logic and Neural Net Theory. *Proceedings of the 1990 Workshop on Fuzzy Logic and Neural Networks* , NASA/JSC.
14. Lee, Chuen-Chien: A Self-Learning Rule-Based Controller Employing Approximate Reasoning and Neural Network Concepts. *International Journal of Intelligent Systems*, vol. 5, 1990.
15. Lee, Chuen-Chien: Modeling behavioral substrates of associative learning and memory: adaptive neural models. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, 1990.

54-38
N91-27 107,7

P-15

EVALUATING THE EFFECT OF ACCURACY RATIOS
ON THE PERCENT OF CALIBRATIONS
WHICH ARE OUT OF TOLERANCE

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

VO 181760

Prepared by: Sharon E. Navard, Ph.D.
Academic Rank: Assistant Professor
University & Department: Virginia Commonwealth University
Department of Mathematical Sciences
Richmond, Virginia 23284-2014

NASA/JSC

Directorate: Safety, Reliability and Quality Assurance
Division: Quality Assurance and Engineering
Branch: Flight Systems Quality Engineering
JSC Colleague: Otho T. Dickerson
Date Submitted: August 10, 1990
Contract Number: NGT 44-005-803

ABSTRACT

The standard practice in calibration laboratories across the country, including the Measurement Standards and Calibration Laboratory (MSCL) at the Johnson Space Center, is to use accuracy ratios to determine if instruments are in-tolerance rather than computing the actual uncertainty associated with the instruments. In the past the accepted practice was to use an accuracy ratio of 10:1, but then state-of-the-art advanced to the point where the 10:1 ratio could no longer be maintained, and the ratio was arbitrarily lowered to 4:1. It is now becoming increasingly difficult to maintain the 4:1 accuracy ratio, and in some cases 1:1 is the best that can be achieved. However, the effect of using these small accuracy ratios on the number of mistakes made in classifying instruments as in- or out-of-tolerance is completely unknown.

In order to assess the effect of using accuracy ratios in calibration, a simulation program was written to compute the proportion of instruments determined to be out-of-tolerance which were actually in, denoted by α , and the proportion of instruments determined to be in-tolerance which were actually out, denoted by β . This was done for accuracy ratios of 1:1 to 10:1, for one to five progressive calibrations, under varying standard and instrument conditions. Selected results are presented and explained in this report; the full set of results, as well as the simulation program itself, can be obtained from the author.

INTRODUCTION

The Measurement Standards and Calibration Laboratory (MSCL) at the Johnson Space Center is responsible for calibrating all of the instruments used at JSC and by offsite contractors. Calibration involves determining if an instrument measures accurately within its tolerance as specified by the manufacturer and, if it does not, adjusting or repairing it so that it does. The time interval between calibrations for each instrument is determined so that, in theory, it will be re-calibrated before it drifts far out of tolerance.

Each calibration that is performed must possess a property known as *traceability*. This means that the standard which was used to do the calibration must be traceable back to either an intrinsic standard or to the national standards maintained by the National Institute of Standards and Technology (NIST). Intrinsic standards are those which are based on physical laws of nature or naturally occurring phenomena, such as the speed of light or the triple point of water. These standards are created or maintained within the Reference Standards Laboratory at JSC, and no interaction with the NIST is necessary for these calibrations. For those quantities for which no intrinsic standard is known, however, the national standards maintained by the NIST are the legal basis for a measurement system in the United States.

While the intrinsic and national standards are by definition the "true" values, any measurements made on them are, unfortunately, subject to error. In order for a calibration to be meaningful, the magnitude of the possible error must be known. There are two characteristics of the measurements which are of interest: the *accuracy*, which determines how close the measurements are, on average, to the true value, and the *precision*, which determines how close repeated measurements made with the same instrument are to each other. An instrument can be accurate but imprecise, and vice versa.

There are several levels of accuracy and precision of standards used in the MSCL. The most accurate are the *reference standards*, which are the most precise standards available, and are calibrated either in the MSCL directly from the intrinsic standards or by the NIST directly from the national standards. The reference standards are recalibrated periodically to maintain their level of accuracy, a process which can be very expensive and time consuming, especially if it involves physically returning the standards to the NIST. Thus, the integrity of the reference standards is closely guarded. Since merely handling the standards can reduce their accuracy, the reference standards are used only to calibrate the *transfer standards*, which in turn are used only to calibrate the *working standards*. It is the working standards which are actually used to calibrate customers'

instruments. They are periodically recalibrated by the transfer standards, which are in turn recalibrated by the reference standards. Thus the accuracy of the reference standards is maintained as long as possible.

The problem with using this progression of standards is that accuracy is lost at each successive step. When a reference standard is used to calibrate a transfer standard, the measurements given by the reference standard are taken to be the actual values; but due to the inaccuracy and imprecision of the reference standard, these values will differ from the "true" values, thus making the reference standard inaccurate by this amount. Also, the imprecision of the transfer standard itself contributes to the inaccuracy, increasing its magnitude. When the transfer standard is then used to calibrate the working standard, its measurements are taken to be "true," and hence this larger inaccuracy is passed on, and is in turn increased by the imprecision of the transfer and working standards.

The result of this is that, when calibrating an instrument which has a certain precision as stated by the manufacturer, the calibrating standard must be even more precise. The precision of an instrument is stated as a *tolerance*, which is theoretically the largest possible magnitude of the difference between the measured and "true" values, assuming the instrument is accurate. The instrument is said to be calibrated, or *in-tolerance*, if a very high percentage (usually more than 99%) of measurements that it gives are within its stated tolerance of the "true" value. Thus to determine if an instrument is calibrated, it is necessary to know the precision and accuracy of the standard; and since the "true" value is unknown, mistakes will be made. The aim of this report is to determine how many mistakes are made under certain circumstances.

DETERMINING THE TOLERANCE OF AN INSTRUMENT

Determining the precision and accuracy of an instrument, and thus the uncertainty associated with it, can be a very difficult and time consuming problem. Extensive literature has been devoted to the subject--see, for example, Abernethy and Benedict (1985); Cameron (1976); Colclough (1987); Croarkin (1984); Eisenhart (1963); Ku (1966); Ku and Judish (1986); and Schumacher (1981). Unfortunately, due to the time involved in computing uncertainties, the common practice in calibration laboratories across the country is to employ the use of accuracy ratios rather than actually computing the uncertainty. A brief description of both methods follows.

Random and Systematic Errors

There are two types of errors which must be evaluated in order to compute the uncertainty of an instrument--the *random error*, which is a measure of precision, and the *systematic error*, which is a measure of the accuracy or offset. The total uncertainty, or tolerance, of the instrument is then the sum of the random and systematic errors.

Since the random error is a measure of precision, it can be estimated by taking repeated measurements with the instrument, under all of the different operational and environmental conditions with which the instrument will be used. The random error is then taken to be some multiple of the standard deviation of the measurements. When the manufacturer specifies a tolerance for an instrument, this tolerance is an estimate of the random error of the instrument; for most of the instruments used in the MSCL, the tolerance is set at three standard deviations. When there is more than one source of random error, such as in progressive calibrations or very complex calibrations, the random errors from the different sources can be combined in quadrature.

The systematic error is much more difficult to compute because, unlike the random error, it depends on the unknown "true" value and cannot be ascertained by repeated measurements. One method of estimating the systematic error is to use the tolerance of the standard used to do the calibration. When there is more than one source of systematic error, there is no general agreement on a method of combining them. Because they are not random, they cannot necessarily be expected to cancel each other out, so the most conservative method of combining them is a direct sum. However, some sources (see, for example, Schumacher (1988)) argue that since systematic errors from different sources are independent of each other, they can be viewed as random observations from another process and can thus be treated as standard deviations and combined in quadrature.

The ideal situation in doing progressive calibrations would be to compute the random and systematic errors at each step and keep track of them, combining them at each successive calibration. If this were done, the MSCL could report to each customer the actual uncertainty associated with his instrument, and also would be able to determine what proportion of instruments will be classified incorrectly as being in- or out-of-tolerance. Unfortunately, due to the time involved with computing random errors, this is not done; instead, accuracy ratios are employed to determine whether or not the instrument reads accurately within the manufacturer's specifications.

Accuracy Ratios

Calibration by using accuracy ratios is much faster and simpler than the method described above because no repeated measurements are taken and the overall uncertainty is not computed. If a test instrument with tolerance T is to be calibrated using an accuracy ratio of $R:1$, then the standard used to calibrate it must have a tolerance of T/R . A quantity is then measured using the standard, and the resulting reading is taken to be the "true value." The same quantity is then measured using the test instrument, and if the reading is within T of the assumed "true value," the test instrument is said to be in-tolerance; otherwise, it is said to be out-of-tolerance. Any time the reading on the test instrument is more than $(0.7)T$ from the assumed true value, the test instrument is adjusted to read the same as the standard.

Since this method of calibration is much faster than the other, it is the method that is used in the calibration labs of all of the NASA centers as well as other calibration labs throughout the country. The JSC Metrology Requirements Manual (1990) specifically advocates the use of a 4:1 accuracy ratio. The reason for using it is expediency; the vast majority of the theory of calibration in the literature is about computing uncertainties. Very little work has been done on determining how well the use of accuracy ratios actually works. It is undesirable from a theoretical point of view because it loses all information about the actual uncertainty of the instrument. Furthermore, the proportion of mistakes made in determining if an instrument is in- or out-of-tolerance is unknown.

In the past, no one worried about what was lost by using accuracy ratios because the accepted practice was to use a 10:1 ratio, which intuitively seemed to guarantee accurate calibrations. However, as the technology improved, customers were able to obtain more precise instruments and it became impossible to maintain the 10:1 ratio. At this point, about twenty years ago, the accuracy ratio was arbitrarily lowered to 4:1 by tacit agreement among the calibration community. The effect of lowering the ratio was unknown, but the consensus seemed to be that it still did a good job and still no one worried about it; theoreticians continued to ignore what was being done in practice and continued to research computing uncertainties. However, the state-of-the-art has now improved to the point where the 4:1 ratio can no longer be maintained in many disciplines and pressure is being felt in the calibration community to lower the accuracy ratio to 3:1; the U. S. Navy has already done so. Even this won't solve the problem in some disciplines, however, where the best accuracy ratio that can be maintained is 1:1. Metrologists are finally starting to worry.

Because the calibration community has only recently become aware that it has a problem, very little work has been done on determining the effect of using accuracy ratios on the proportion of instruments which are incorrectly determined to be in- or out-of-tolerance. Nothing on the subject has been published in the major statistical journals, although some papers have recently been presented on the problem at metrology conferences. See, for example, Capell (1988); Capell (1989); and Schumacher (1988). Of these, only Schumacher addresses the specific problem of computing the proportion of instruments which are incorrectly classified. He actually computes the proportion of products which are incorrectly determined to meet or not meet specifications when the measuring instrument has a tolerance that is a fraction of the specifications, but this is equivalent to the case of using an accurate standard to calibrate a customer's instrument.

THE QUESTION

The specific question to be addressed is this: What proportion of instruments are incorrectly determined to be in- or out-of-tolerance, after one to five progressive calibrations, using accuracy ratios of 1:1 to 10:1? This should be answered for different proportions of the instruments being calibrated being actually out-of-tolerance, and for the standards being both in-tolerance and out-of-tolerance.

Notation

The following notation will be used in the remainder of this report.

- α = proportion of instruments determined to be out-of-tolerance which are actually in tolerance
- β = proportion of instruments determined to be in-tolerance which are actually out-of-tolerance
- I_0 = the reference standard
- I_i = the i^{th} instrument progressively calibrated after the reference standard
- E_i = the error distribution of I_i
- k = the number of standard deviations of E_i defining "in-tolerance"
- σ = standard deviation of reference standard error distribution, E_0

100p = percent that calibrating standard is out-of-tolerance
 r:1 = accuracy ratio

Employing this notation, and assuming that the accuracy ratios are exactly correct, the reference standard I_0 will have tolerance $k\sigma$, while I_i will have tolerance $r^i k\sigma$, $i = 1, 2, \dots, 5$. If I_i is in-tolerance when it is used as a standard, $i = 0, \dots, 4$, then the standard deviation of the error distribution E_i is $r^i \sigma$; otherwise it is $(1 + p)r^i \sigma$.

Assumptions of the Model

In order to compute α and β , it is first necessary to define the specific probability model that is being observed. When working with measurement errors, a standard assumption to make is that the errors have a normal distribution with mean zero. It is also assumed that an r:1 accuracy ratio is exactly maintained throughout the progression of calibrations. Therefore,

$$E_i \sim N(0, [r^i \sigma]^2), i = 0, \dots, 5.$$

If I_i is actually 100p% out-of-tolerance, then

$$E_i \sim N(0, [(1 + p)r^i \sigma]^2).$$

Note that if $p = 0$, then this distribution reduces to that given when the standard is in-tolerance. Henceforward, the general form of the distribution of E_i will be used.

It is necessary to know not only the error distribution, but also the distribution of the measurements themselves. With E_i defined as above, the distribution of measurements X_i taken from I_i will also have a normal distribution with some mean μ_i and the same standard deviation; that is,

$$X_i \sim N(\mu_i, [(1 + p)r^i \sigma]^2), i = 0, \dots, 5.$$

Thus far the assumptions have been very straightforward, but at this point one is faced with a dilemma: what is μ_i ? Ideally, μ_i will be equal to the "true" value at each step, which would mean that the instruments are all exactly accurate. Unfortunately, this will almost certainly not be the case; a systematic error will probably be present. But what is the magnitude of the systematic error? The metrologists at the MSCL were unable to give any practical insight into this problem; systematic errors are the ones which are nearly impossible

to estimate, even if the time and resources to do so are available. It is thus necessary to use some intuitive reasoning.

The reference standard I_0 must be treated differently from the others because it is calibrated using actually *known* values, that is, the intrinsic or national standards. For simplicity's sake, denote the "true" value by μ . When I_0 is being calibrated, it should be made to register exactly μ . However, due to the imprecision of the instrument itself, it will be set to read a value slightly different from μ . Since $E_0 \sim N(0, [(1 + p)\sigma]^2)$, μ_0 will on the average be equal to μ , but will vary from it an amount determined by the distribution of E_0 . Thus,

$$\mu_0 \sim N(\mu, [(1 + p)\sigma]^2).$$

The exact value of μ_0 will depend on the particular observations made by the instrument at the time of the calibration. In practice, calibration of the reference standard is done at the NIST, if national standards are used, or in the Reference Standards Lab of the MSCL, if intrinsic standards are used. In either case, repeated measurements are made, and σ is then estimated. At this step, accuracy ratios are not used.

Determination of μ_i , $i = 1, \dots, 5$, requires further reasoning. The situation is not as simple as that of determining μ_0 because, instead of actually setting μ_i based on observations from I_i , it is now necessary to know what μ_i is before any observations are ever taken. This is because of the method associated with using accuracy ratios--one observation will be taken from the standard I_{i-1} and the test instrument I_i , and these observations will then be compared to determine if I_i is in-tolerance.

At this point, it is necessary to make some assumptions, which may and may not be correct. First, assume that I_i is, on the average, in-tolerance. However, a certain percentage of the instruments being calibrated are found to be out-of-tolerance; at the MSCL, this percentage is 10% or more. Thus the possible values of μ_i must be allowed to shift about μ in either direction, out to a distance that is determined by the percent of instruments coming in which are actually out of tolerance. For simplicity, it will again be assumed that μ_i has a normal distribution with mean μ , but now the standard deviation must be larger than that of E_i . In fact, it is assumed that

$$\mu_i \sim N(\mu, [D(1 + p)r^i\sigma]^2),$$

where D is the drift factor that determines what percent of the instruments coming in to be calibrated are actually out of tolerance.

Similar to the case of μ_0 , the exact value of μ_1 will depend on the particular instrument which is being calibrated.

Defining the measurement distributions as above, the values of α and β will vary depending on the values of p and D . However, due to the nature of the normal distribution, the particular values of μ and σ are unimportant; the probabilities will be the same no matter what values are used. Therefore it is assumed without loss of generality that $\mu = 0$ and $\sigma = .01$.

THE EFFECT OF USING ACCURACY RATIOS

Now that the specific model has been defined, it is now possible to compute α and β . Unfortunately, it is possible to obtain an analytical solution only in a limited situation--when both the standard and test instrument are accurate. Even then it is only possible to obtain the solution for one calibration, because when this calibration is performed a systematic error is introduced. Therefore, in order to evaluate α and β , it is necessary to perform a simulation.

Simulation Program

The simulation program was written in SAS/IML, the Interactive Matrix Language of the statistical package SAS. The random number generator used was the SAS function RANNOR. The parameter values used were:

$k = 2.4, 3;$

$p = 0, .1, .25, .5, 1, 2, 3;$ and

$D = 15\%, 32\%, 65\%, 92\%.$

For each combination of these parameters, 10,000 iterations were performed using accuracy ratios of 1:1 to 10:1, and progressing 5 steps down from the reference standard.

The algorithm for the program is as follows. At each iteration, for $i = 1$ to 5, observations x_{i-1} and x_i are obtained from I_{i-1} and I_i by first generating the means, and then the observations themselves. Once the mean is generated for an instrument when it is the test instrument, the same mean is used at the next step when that same instrument becomes the standard. Since the "true" value is zero, I_i is defined to be *actually* in-tolerance if $|x_i| < kr^i\sigma$, where r is the accuracy ratio. Since the perceived true value is x_{i-1} , I_i is *perceived* to be in-tolerance if $|x_{i-1} - x_i| < kr^i\sigma$. There are therefore four possible outcomes: the instrument is actually in-tolerance, and is also perceived to be in-tolerance; the instrument is actually out-of-

tolerance and is also perceived to be out-of-tolerance; the instrument is actually in-tolerance but is perceived to be out-of-tolerance (an α error); or the instrument is actually out-of-tolerance but is perceived to be in-tolerance (a β error). The outcome for each step is observed, and the total number occurring of each outcome out of the 10,000 iterations is counted. In keeping with the procedure actually practiced in the MSCL, if $|x_{i-1} - x_i| > (.7)kr^1\sigma$, I_i is adjusted to read the same as I_{i-1} . This procedure is done for accuracy ratios of 1:1 to 10:1.

Results

The output of the program for selected parameter values is shown in Tables 1 - 4; output using the other parameter values follows a similar pattern. In general, it can be seen that, as one would expect, no matter what the parameter values are, α and β decrease as the accuracy ratio is increased. For a particular accuracy ratio, α is generally larger than β , with the magnitude of the difference decreasing as the accuracy ratio is increased. This is desirable because α is a less severe error than β ; it is also the same pattern that was observed in Schumacher (1988). When the accuracy ratio is 1:1, α and β both increase as progressive calibrations are made from one step to the next, but this increase vanishes as the accuracy ratio is increased. For accuracy ratios of 2:1 and 3:1, there is an increase in α and β from the first step to the second, but they remain constant afterward; for accuracy ratios of 4:1 and higher, α and β remain fairly constant for all five steps down.

Comparing Tables 1 and 2, it can be seen that as "in-tolerance" is changed from 3σ to 2.4σ , the percent of errors made increases. This is to be expected, because errors are more likely to be made when the observation is near the edge of the tolerance range, and when the tolerance range is reduced, a larger proportion of observations will be near the cutoff points. Comparing Tables 1 and 3, it can be seen that, as one would expect, if the standard is out-of-tolerance, more mistakes will be made. The farther out-of-tolerance the standard is, the more mistakes will be made. Finally, comparing Tables 1 and 4, one can see that as more instruments are out-of-tolerance coming in, fewer mistakes are made. While this may seem surprising at first glance, it makes sense because the way the model was defined, fewer observations are near the edge of the tolerance range. The farther the instruments drift out-of-tolerance, the easier it is to detect, and fewer mistakes will be made.

TABLE 1.- α AND β WHEN TOLERANCE = 3σ , STANDARD IS IN-TOLERANCE, AND 15% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

Step	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
	α	β	α	β	α	β	α	β	α	β
1	.131	.043	.050	.026	.030	.022	.023	.017	.018	.015
2	.210	.052	.063	.033	.037	.023	.026	.016	.020	.014
3	.257	.054	.063	.031	.036	.025	.025	.017	.020	.015
4	.286	.056	.068	.033	.036	.022	.029	.022	.021	.013
5	.322	.058	.065	.034	.036	.022	.026	.016	.020	.017

Step	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
	α	β	α	β	α	β	α	β	α	β
1	.014	.013	.011	.010	.01	.008	.010	.008	.012	.007
2	.015	.016	.013	.011	.011	.009	.010	.009	.009	.009
3	.016	.012	.013	.011	.012	.009	.008	.010	.009	.007
4	.018	.012	.013	.010	.012	.010	.010	.009	.008	.010
5	.017	.013	.013	.010	.012	.010	.010	.009	.010	.007

TABLE 2.- α AND β WHEN TOLERANCE = 2.4σ , STANDARD IS IN-TOLERANCE, AND 15% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

Step	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
	α	β	α	β	α	β	α	β	α	β
1	.167	.055	.066	.035	.041	.027	.030	.022	.022	.017
2	.246	.058	.085	.039	.050	.031	.033	.021	.023	.018
3	.304	.067	.084	.038	.045	.030	.034	.021	.027	.019
4	.342	.059	.084	.041	.047	.027	.029	.020	.025	.019
5	.365	.055	.084	.038	.048	.029	.033	.022	.024	.016

Step	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
	α	β	α	β	α	β	α	β	α	β
1	.020	.016	.016	.012	.017	.01	.012	.012	.011	.009
2	.019	.016	.016	.011	.016	.013	.011	.011	.012	.009
3	.020	.015	.019	.015	.015	.011	.014	.012	.011	.011
4	.019	.014	.017	.012	.014	.016	.015	.012	.011	.01
5	.017	.016	.013	.014	.016	.011	.014	.012	.011	.010

TABLE 3.- α AND β WHEN TOLERANCE = 3σ , STANDARD IS 50% OUT-OF-TOLERANCE, AND 15% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

Step	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
	α	β	α	β	α	β	α	β	α	β
1	.171	.048	.074	.036	.040	.023	.028	.019	.024	.016
2	.262	.052	.081	.035	.047	.025	.033	.022	.028	.016
3	.319	.055	.083	.037	.045	.030	.034	.019	.026	.019
4	.364	.052	.087	.036	.047	.028	.036	.018	.025	.020
5	.394	.054	.085	.035	.044	.028	.034	.024	.026	.018

Step	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
	α	β	α	β	α	β	α	β	α	β
1	.018	.013	.014	.014	.013	.011	.011	.009	.013	.008
2	.022	.017	.016	.013	.013	.012	.014	.011	.014	.008
3	.020	.014	.017	.013	.016	.012	.012	.012	.011	.008
4	.022	.015	.019	.011	.016	.012	.013	.011	.011	.009
5	.020	.016	.016	.014	.016	.012	.011	.01	.011	.011

TABLE 4.- α AND β WHEN TOLERANCE = 3σ , STANDARD IS IN-TOLERANCE, AND 32% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

Step	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
	α	β	α	β	α	β	α	β	α	β
1	.110	.067	.047	.039	.037	.028	.023	.022	.019	.016
2	.180	.084	.067	.052	.039	.031	.029	.024	.022	.020
3	.212	.099	.066	.050	.038	.031	.030	.022	.021	.017
4	.247	.099	.068	.050	.037	.033	.029	.025	.021	.017
5	.273	.102	.060	.046	.038	.036	.029	.024	.023	.021

Step	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
	α	β	α	β	α	β	α	β	α	β
1	.015	.014	.014	.013	.013	.013	.011	.010	.010	.010
2	.019	.018	.017	.013	.012	.011	.012	.013	.011	.011
3	.016	.018	.016	.013	.015	.013	.010	.013	.011	.010
4	.019	.016	.017	.016	.012	.011	.010	.010	.011	.011
5	.018	.017	.016	.016	.015	.014	.011	.011	.010	.011

CONCLUSIONS

When using the results of this simulation to determine the effect of accuracy ratios on the percent of instruments which are incorrectly determined to be in- or out-of-tolerance, it is necessary to remember that the results obtained apply only to the specific model that was used. While most of the assumptions seem fairly reasonable, the method of determining the means of the instrument measurement distributions may not closely approximate what is actually happening in practice. Furthermore, the definition of "in-tolerance" when using accuracy ratios is not the same as the standard definition. When using accuracy ratios, the instrument is said to be in-tolerance if the one observation taken is within a certain tolerance of the "true" value. However, the one observation taken may be an extreme value from the measurement distribution, and in fact most observations taken from that instrument would be outside of the tolerance. This possibility was not considered in this program.

With this in mind, it is now possible to answer the questions that were asked. First of all, the effect of lowering the accuracy ratio from 4:1 to 3:1 is, in almost all cases, an increase of approximately 0.01 in both α and β . While this may sound like a small price to pay, if one considers the percent increase in the number of errors made, it ranges from approximately 25% to 50%, depending on the parameter values--a substantial increase. Secondly, using a 1:1 accuracy ratio is substantially worse, and the magnitude of the errors increases as the progressive calibrations are made. The size of the errors made using the 1:1 ratio is almost certainly unacceptable in any situation.

Nevertheless, situations do exist where accuracy ratios of 1:1 to 3:1 are the best that can be maintained. What can be done in these situations? One suggestion is to abandon accuracy ratios altogether and keep track of the systematic and random errors at each successive calibration. The metrologists will immediately counter this suggestion with the fact that they have neither the time nor the personnel to do this for every instrument they calibrate.

Perhaps a compromise can be reached. Since the transfer and working standards are not calibrated very often, it may be practical to take the time to make repeated observations on them to obtain an accurate estimate of their uncertainties. This might lead to a smaller tolerance for the working standards than that obtained by using the accuracy ratios, which would in turn make larger accuracy ratios possible for calibrating customers' instruments. If the tolerance gets bigger instead of smaller, this would demonstrate that the problem with calibration is even more serious than previously believed. In either case, the MSCL would have a better idea of how accurate its calibrations are than is now known.

REFERENCES

- Abernethy, R. B.; and Benedict, R. P.: Measurement Uncertainty: A Standard Methodology. *ISA Transactions*, vol. 24, no.1, 1985, pp. 75-79.
- Cameron, Joseph M.: Measurement Assurance. *Journal of Quality Technology*, vol. 8, no. 1, Jan. 1986, pp. 53-55.
- Capell, Frank: From 4:1 to SPC. Paper presented at the National Conference of Standards Laboratories 1988 Workshop and Symposium (Washington, D. C.), Aug. 14 - 18, 1988.
- Capell, Frank: More on 'From 4:1 to SPC.' *National Conference of Standards Laboratories 1989 Workshop and Symposium Technical Presentations*, July 9-13, 1989, pp. 8-1 - 8-15.
- Colclough, A. R.: Two Theories of Experimental Error. *Journal of Research of the National Bureau of Standards*, vol. 72, no. 3, May-June 1987, pp. 167-185.
- Croarkin, Carroll: Measurement Assurance Programs Part II: Development and Implementation. National Bureau of Standards Special Publication 676-II, Apr. 1984.
- Eisenhart, Churchill: Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems. *Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation*, vol. 67C, no. 2, April-June 1963, pp. 161-187.
- JSC Metrology Requirements Manual. JSC M 8070C, April 1990.
- Ku, Harry H.: Notes on the Use of Propagation of Error Formulas. *Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation*, vol. 70C, no. 4, Oct.-Dec. 1966, pp. 263-273.
- Ku, Harry H.; and Judish, Robert M.: Fundamentals of Error Analysis. *Proceedings of the IEEE*, vol. 74, no. 1, Jan. 1986, pp. 25-27.
- Schumacher, Rolf B. F.: Systematic Measurement Errors. *Journal of Quality Technology*, vol. 13, 1981, pp. 10-24.
- Schumacher, Rolf B. F.: MIL-STD 45662: Conflicting Positions on Out-of-Tolerance Measurements. *Proceedings of the 1988 Measurement Science Conference*, Jan. 1988, pp.99-105.

55-12
20220
N91-271089

Developing a Taxonomy for Mission Architecture Definition

Final Report

NASA/ASEE Summer Faculty Fellowship Program— 1990

Johnson Space Center

42086788

Prepared By:	Deborah J. Neubek
Academic Rank:	Visiting Assistant Professor
University & Department:	University of Houston College of Architecture Houston, TX 77204-4431
NASA/JSC	
Directorate:	Lunar & Mars Exploration Program Office
Division:	System Engineering & Integration
JSC Colleague:	Norman H. Chaffee
Date Submitted:	9-17-90
Contract Number:	NGT-44-005-803

Abstract

The Lunar & Mars Exploration Program Office (LMEPO) was tasked to define candidate architectures for the Space Exploration Initiative to submit to NASA senior management and an externally constituted Outreach Synthesis Group. A systematic, structured process for developing, characterizing and describing the alternate mission architectures, and applying this process to future studies was developed. The work was done in two phases: first, National Needs were identified and categorized into objectives achievable by the Space Exploration Initiative. Secondly, a Program Development Process was created which both hierarchically and iteratively describes the program planning process.

Introduction

The Space Exploration Initiative (SEI) is the program being developed in response to President George Bush's proposal to take the U.S. back to the Moon then on to Mars. Whereas in 1961 it took a crisis (the space race) to galvanize the nation to reach for the stars, SEI is based on the President's belief that space holds an opportunity for greatness.

After the initial speech given by the President on July 20, 1989, NASA undertook a 90-Day Study which resulted in a potential approach to achieving these goals. The National Space Council then decided to broaden the input to this activity and requested that an independent committee synthesize data collected from an unrestricted range of public and private institutions and individuals (Outreach). This committee, termed the Synthesis Group and chaired by General Thomas Stafford, is due to produce a report in March, 1991 which suggests "at least two" alternatives for embarking on the Space Exploration Initiative.

Determining Program Objectives from National Needs

To fully reap the benefits of SEI, it is important to identify the National Needs which can be met by the program and a process by

which links may be forged. Linking National Needs to program planning results in two separate but related products:

- Program drivers (key goals and objectives to be achieved) which are INPUTS to the program planning process.
- Program benefits which are OUTPUTS from doing the program and are what the American public wants to receive from their investment.

It was in this context that a team of Level I and Level II people performed an analysis of national needs and policies which resulted in the chart seen in figure 1. This chart lists National Needs in terms of achievable goals and objectives which might guide program planning. The group identified seven primary areas of National Needs which may be addressed by the SEI. These are to: Invigorate National Spirit and Pride; Expand Human Knowledge; Stimulate Education; Strengthen the Economy; Improve the Quality of Life; Improve International Relations; and Strengthen National Security.

Under each of these areas, specific objectives were identified which might be set by a policymaker as an area of emphasis. For example, if "Enhance and improve potential career opportunities" were designated as a high priority objective, the SEI architecture developed might show additional coop programs included in the management plan.

It is important to note that the SEI program will meet all the objectives listed to some extent. However, the set of objectives selected as program drivers will influence the final implementation of the program. An alternate set of objectives would yield a program which is different in its make-up. Therefore, different architectures may be created which all meet the basic National Needs, however, architecture "A" may have enhanced commercialization programs whereas architecture "B" might place more of an emphasis on scientific research.

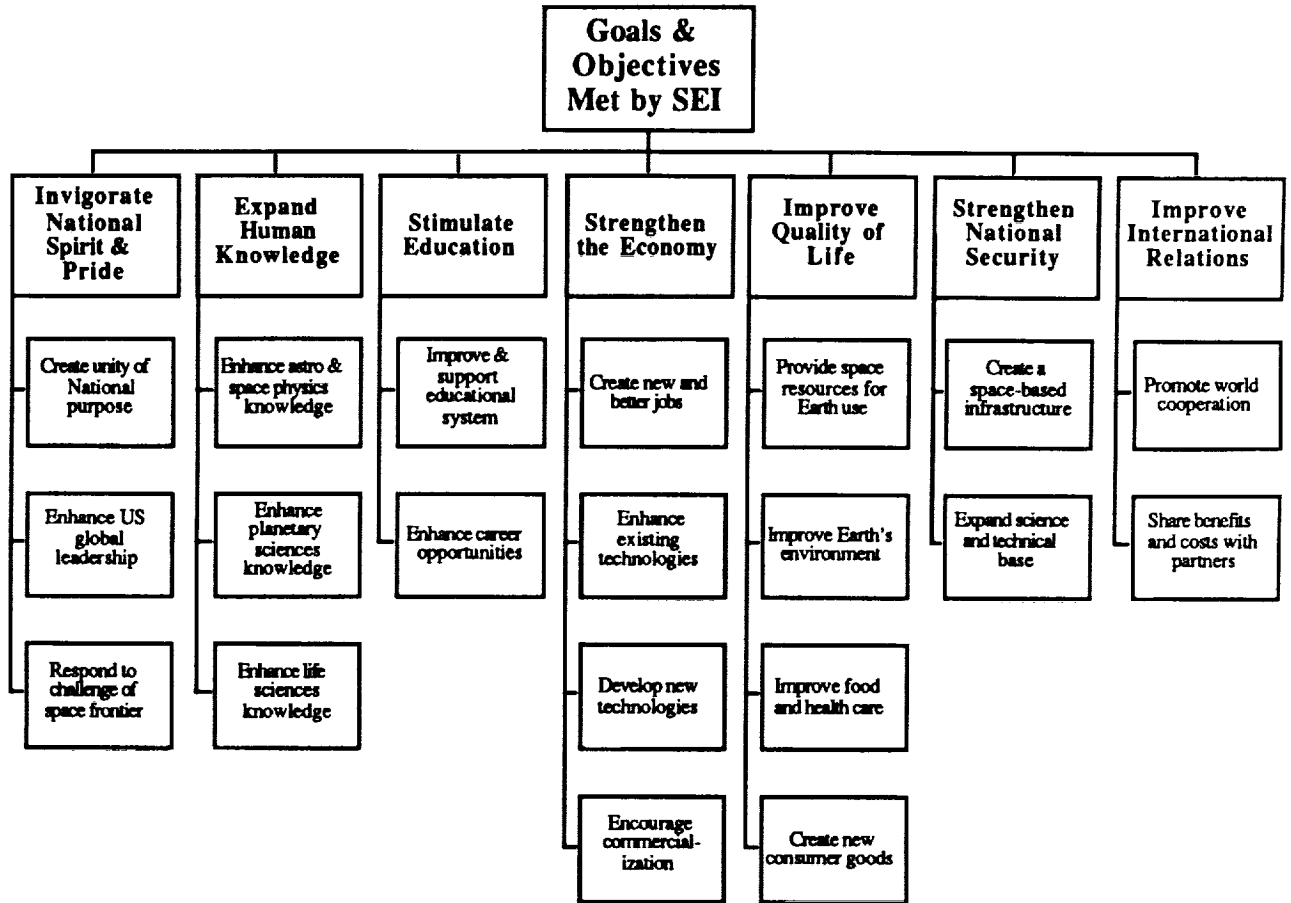


Figure 1: National Needs Stated as Policy Objectives

Developing a Process to Link National Needs to Implementations

Once the National Needs were identified, it was necessary to develop a process by which they could be physically linked to specific program goals and features. In the most basic of terms, three questions must be answered to develop a program: Why go? What will we do? How will we do it? The chart in figure 2 illustrates a top-down, hierarchical approach for linking National Needs to program implementation. This chart serves as a logical outline which shows the different levels of definition needed to complete the description of a program architecture.

Developing a program as complex as the SEI is not strictly hierarchical, however. It is an iterative process with many inter-related steps. Figure 3 illustrates the primary steps and their relationships.

The first step in developing a program is to define the **program purpose** by determining the primary drivers: Why are we going? What do we want to accomplish?

These questions may be answered through:

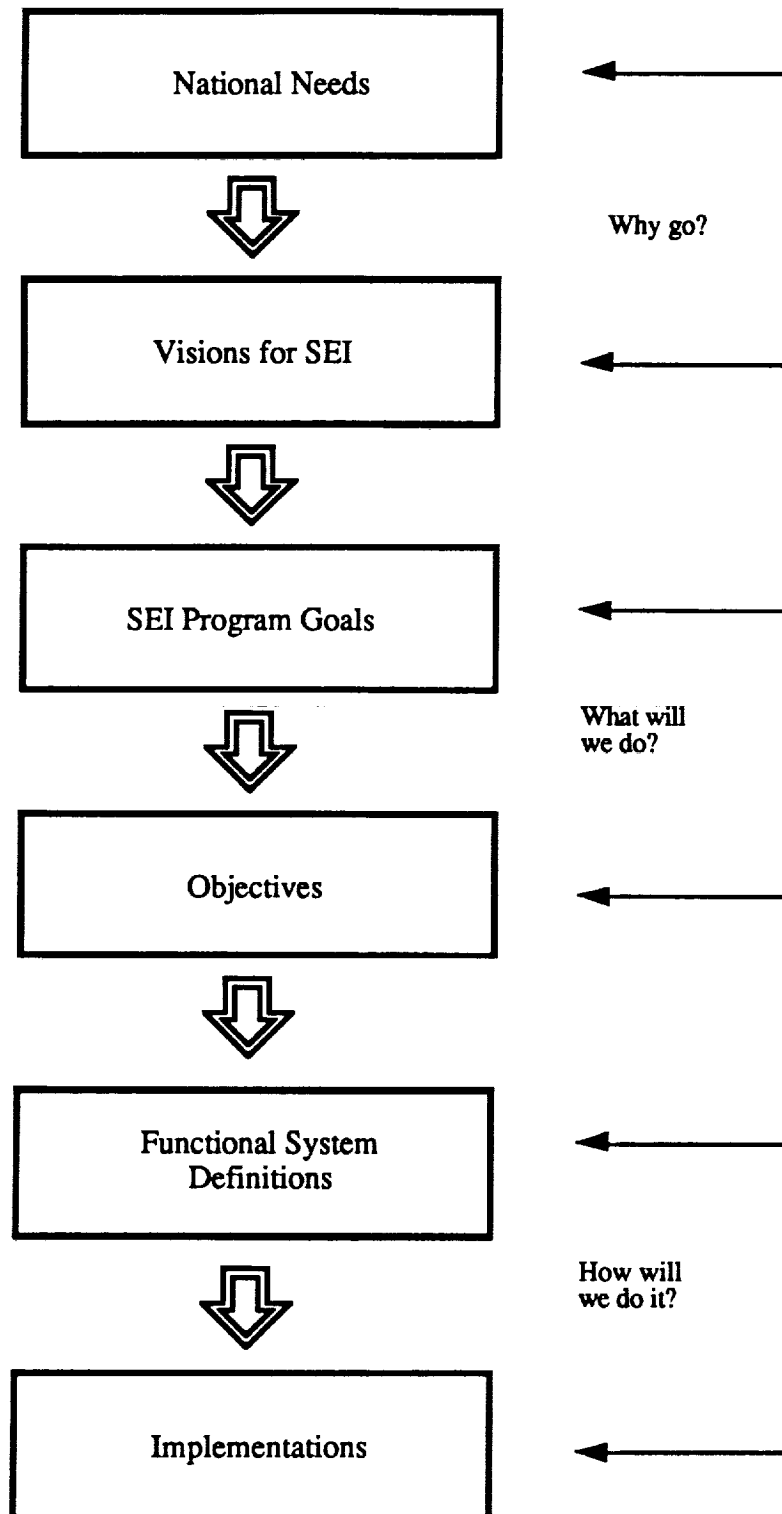
- understanding the overarching National Needs
- establishing Visions which will guide SEI planning and respond to National Needs
- setting Program Goals which define what must be achieved to fulfill the Visions and
- creating a Program Mission Statement which captures the essence of the SEI Program Goals.

Each of these areas may be defined as follows:

National Needs are the same as those discussed previously and shown in figure 1.

Six Visions have been enumerated by the Synthesis Group to guide SEI planning. These are:

- Enhance knowledge of the planets including Earth
- Rejuvenate interest in science and technology



LMEPO/Neubek/Ver. 2.0 7/25/90

Figure 2: Linking National Needs to Implementations

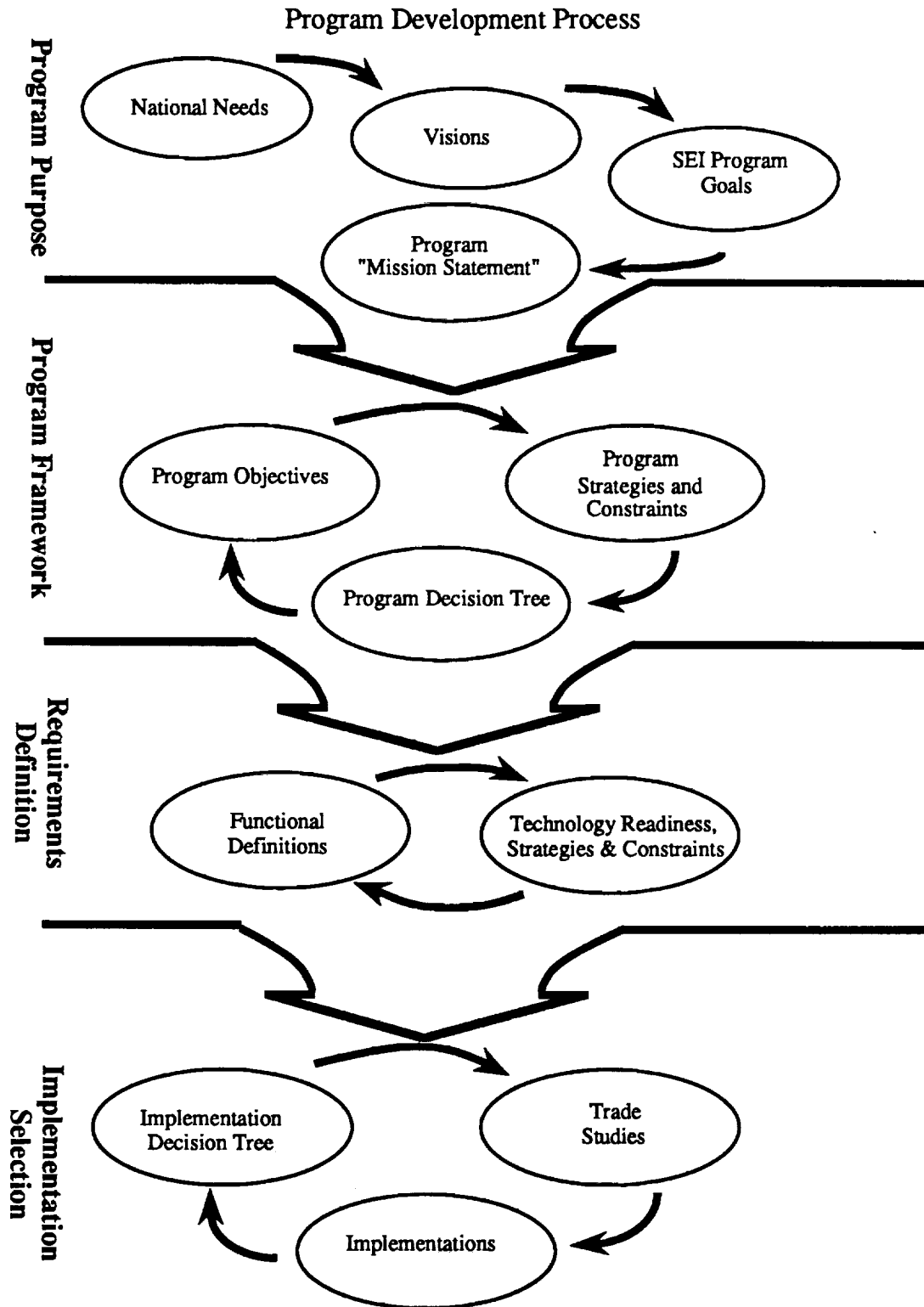


Figure 3: An Iterative Program Development Process

- Refocus the role of U.S. world leadership
- Develop technologies with terrestrial applications
- Facilitate further space exploration and privatization
- Boost the U.S. economy

SEI Program Goals are specific, space-oriented statements which guide and direct the program so the visions are fulfilled. They serve as the "missing link" between "fuzzy" national needs and "hard" technical or programmatic requirements.

A "universal set" of goals, including some which may be incompatible with others may be created. A subset of these goals which are internally consistent may then be selected to create an architectural theme. For clarity it is useful to break these goals into two categories:

- spacefaring ("of or engaged in life in Space"— Webster's)
- administrative ("of or related to policies, procedures or management issues")

The SEI Program Mission Statement captures the essence of the primary drivers for a given program theme. It is made up of a selected subset of SEI Program Goals.

Once the program purpose is established, a **program framework** must be laid out to provide the foundation for requirements definition. This consists of iteratively identifying program objectives, strategies, constraints and key decisions. Program Objectives are those measurable program features which serve to frame what should be done, specifically, to fulfill the program goals. Program Strategies are the internal, long range approaches which cut across systems. They are methods or plans of action for accomplishing the objectives. Program Constraints are those external limitations, guidelines or policies which are imposed upon the program. These may be technical, budgetary or programmatic in nature. Key Decisions are those decision points identified by examining each of the previously selected elements of the Program Purpose and Program Framework and incorporating temporal considerations. A key point in creating the Program Framework is to assure that these decisions are top-level in nature, and deal with programmatic issues.

The next step is to begin **requirements definition**. Systems must be defined in terms of their functional parameters while technology readiness, strategies and constraints must be identified. The Functional Definitions are descriptions which are expressed in terms of capabilities and are grouped into like areas. For example, planetary surface operations is a functional area. Early program decisions may be made through the examination of Technology Readiness, Strategies and Constraints. Once the critical technologies are identified, they can be plotted on a temporal scale and interrelationships drawn. This leads to the creation of a tool which aids in determining which programs have the greatest impact on others, as well as which programs should be started early on in the overall SEI schedule.

Finally **implementation** trade spaces may be selected and studied, leading to decisions for future technology and advanced development programs.

Conclusions

The Program Development Process is a tool by which the SEI program may be planned hierarchically, beginning with National Needs and working down to implementations. A key factor in using it successfully is to identify the appropriate levels of detail and depth which should be addressed at each step. In general, the program should be planned from the top down, defining what should be done, then functionally how it should be done. Only after this has been accomplished can the implementation studies be optimized.

This process is iterative, and not only entails interrelating components at each step, but also assumes the entire process will be repeated as new information is returned from studies. By developing a program for the SEI that is guided by National Needs, SEI can be optimized for returning the greatest benefits possible to the nation.

56-31
N91-27 109-21

DESIGN OF A LUNAR OXYGEN PRODUCTION PLANT

P-12

Final Report
NASA/ASEE Summer Faculty Fellowship Program - 1990

JOHNSON SPACE CENTER

Prepared By: Ramalingam Radhakrishnan, Ph.D., P.E.
Academic Rank: Assistant Professor
University & Department: Prairie View A&M University
Department of Civil Engineering
College of Engineering and Architecture
Prairie View, Texas 77446

PO 782448

NASA/JSC

Division: Solar System Exploration Division
Branch: Space Resources Utilization Office
JSC Colleagues: David S. McKay, Ph.D.
Thomas A. Sullivan, Ph.D.
Date Submitted: August 17, 1990
Contract Number: NGT-44-005-803

ABSTRACT

To achieve permanent human presence and activity on the Moon, oxygen is required for both life support and propulsion. Lunar oxygen production using resources existing on the Moon will reduce or eliminate the need to transport liquid oxygen from earth. In addition, the co-products of oxygen production will provide metals, structural ceramics and other volatile compounds. This will enable development of even greater self-sufficiency as the lunar outpost evolves.

Ilmenite is the most abundant metal-oxide mineral in the lunar regolith. A process involving the reaction of ilmenite with hydrogen at 1000°C to produce water, followed by the electrolysis of this water to provide oxygen and recycle the hydrogen, has been explored.

The objective of this 1990 Summer Faculty Project was to design a lunar oxygen-production plant to provide 5 metric tons (mt) of liquid oxygen per year from lunar soil. The results of this study describe the size and mass of the equipment, the power needs, feedstock quantity and the engineering details of the plant.

INTRODUCTION

The primary objective of this project was to design a plant to produce oxygen on the Moon from lunar materials. The Lunar Oxygen Production Plant is required to produce 5 metric tons (mt) of oxygen per year. The plant process selected for this study is the hydrogen reduction of ilmenite with soil feedstock. The plant consists of equipment for mining, processing, oxygen storage, and hydrogen storage. After construction or emplacement on the Moon, the plant will be teleoperated from Earth. The plant is assumed to have a duty cycle of 90% and a continuous power supply.

A simplified schematic of the hydrogen reduction of ilmenite process is shown in Figure 1. In this process, the bulk raw soil is mined and fed into the plant. The plant separates the very large and fine fractions of the soil so that an acceptable size of the soil can be processed. A system of storage hoppers then feeds the soil containing ilmenite into the top of a three-stage fluidized bed reactor. The reactor heats the soil to 1000°C as the soil flows down the reactor's three stages. Hydrogen gas flows up through the solids, mixes with the particles, and reacts with the ilmenite to form water vapor. An electrolysis cell separates the water vapor into hydrogen and oxygen. The oxygen is liquefied and stored. The hydrogen is fed back into the top section of the reactor. A solids settler removes the reactor residuals (spent solids) from the bottom of the reactor.

This report describes the different plant components. The results of the design are presented. The summary included shows the mass, power, and volume of the plant components.

Assumptions

1. The required production capacity of the plant is 5 mt liquid Oxygen (LOX) per year using bulk lunar soil on the surface.
2. The concentration of ilmenite in lunar regolith can be as high as 10 volume percent (15 weight percent). However, a more likely value of 5 volume percent (7.5 weight percent) will be used in this design. An alternate design based on 10 volume percent of ilmenite (15% wt) concentration in regolith would be similar but the plant would be slightly smaller in size than the current design. Another advantage would be that the plant could produce more oxygen than the required 5 mt capacity if the feedstock contains higher percentage of ilmenite.
3. No beneficiation. This means there will be no enrichment steps included in the plant to increase the concentration of ilmenite in the soil.
4. Mining will be done only during the lunar day. At night, metals in the mining equipment could become brittle unless heated. The hot noon time also will be avoided. This requires the storage of sufficient mined regolith for operation during the night.
5. The plant operation is continuous with a 90 % duty cycle, allowing 10% of the time for trouble-shooting, maintenance and unexpected problems.
6. The process yield (chemical reaction and ilmenite conversion) factor is 0.9.
7. Lunar surface temperature varies from -100°C to +111°C. Assumed feedstock temperature is -20°C.

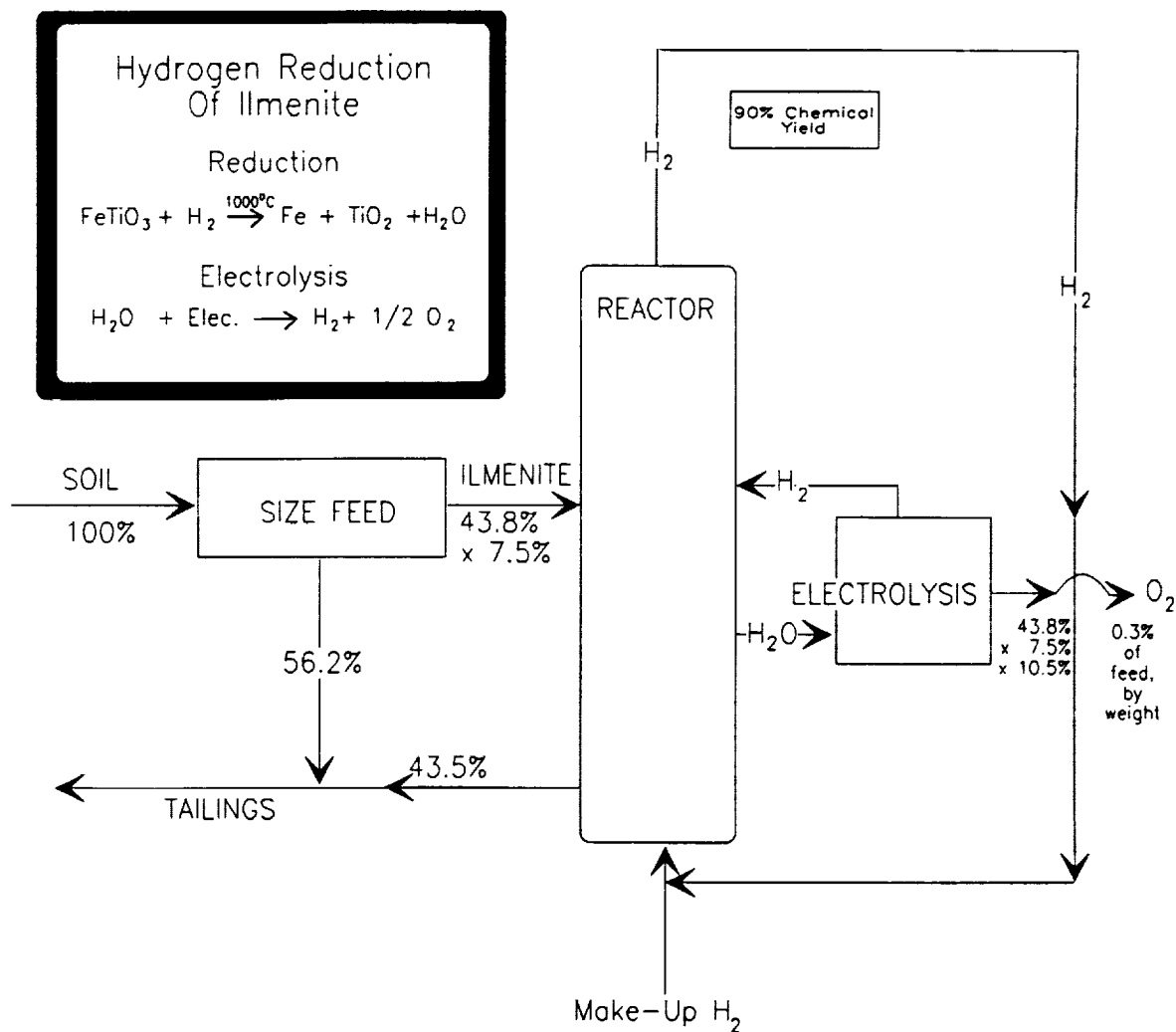


Figure 1
 Simplified Schematic of
 Hydrogen Reduction of Ilmenite Process

LOX PRODUCTION PLANT

The oxygen production process is illustrated in Figure 2. Approximately 1620 metric tons of bulk soil feedstock is required per year. After removing the oversize and undersize rejects, a useful feed of 710 mt is available. As shown in Figure 1, considering a 7.5% wt of ilmenite concentration in the feed and a 90% yield of oxygen during the hydrogen reduction process, LOX production of 5 mt per year is obtained. The important components of the LOX production plant are described in the following sections.

Mining

The mining equipment assumed is a front end loader and hauler. The front end loader (FEL) is a telerobotic vehicle for mining feedstock. The FEL can complete soil mining and will have sufficient time remaining to perform additional base operations.

One hauler is assumed. After being filled by the FEL, the hauler will travel to the process feed bin, dump the load, collect a load of tailings from the discharge bin, dump the tailings at the discharge area, and return to the mining site to start the new cycle over again.

The mining operation is on a 35% duty cycle, equivalent to 3066 hrs. The operation will be discontinued during lunar noon and night. The mining equipment has excess capacity. Since there will be a need for the installation of additional units to increase the oxygen production in the future, the excess capacity of the FEL and hauler is justified.

Bulk Storage Feed Bin

Since the plant is run on a 90% duty cycle and the mining equipment is operated on a 35% duty cycle, the feed bin is sized to hold enough soil to run the plant during the lunar night (14 days). Using high-strength polyethylene fiber Spectra 1000^(TM) in place of Al 2024 is recommended. The shell with 2 mm thick fiber Spectra 1000^(TM), and 1-cm-thick low-density insulation to provide protection from radiation and temperature effects reduces the mass of the bin considerably. Furthermore, this fabric might be compacted into a much smaller volume for transport to the Moon.

Coarse Screen

The coarse Screen removes the oversize size unwanted rocks (material >0.5 mm) in the soil to discard bin (but see recommendations for an alternative).

Fine Vibrating Screen

The minimum allowable feed size to the reactor is 0.45 mm. A vibratory screen was selected to remove fines.

Feed Hoppers (2)

Two hoppers will be used to minimize gas losses from the reactor. The capacity of each hopper is adequate for 3 days of feedstock. Feedstock is fed into the reactor continuously from the bottom hopper which is maintained at 10 atm. pressure. When inventory in the bottom hopper is at an appropriate low point, the top hopper, which is

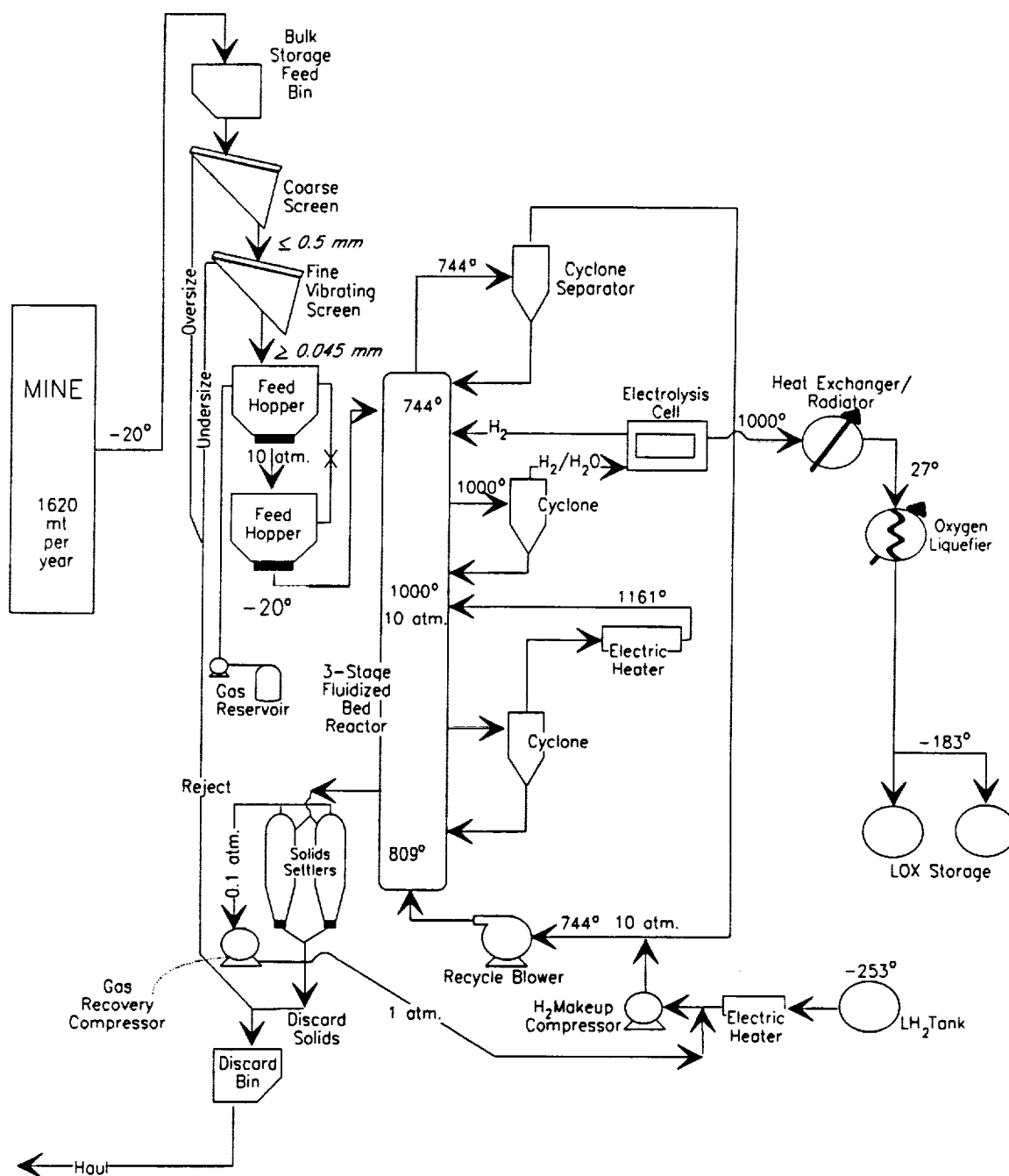


Figure 2
Hydrogen Reduction of Ilmenite
LOX Production Plant

operating normally in vacuum, is pressurized to 10 atm. Then a valve between the two hoppers is opened to rapidly re-inventory the bottom hopper. After the valve between the hoppers is closed, hydrogen gas in the top hopper is recovered. The design allows continuous feed to the reactor and minimizes gas losses from the reactor.

Reactor

Ilmenite reduction takes place in a three-stage fluidized bed reactor (shown in greater detail in Figure 3) at 1000°C and 10 atm. pressure. The reactor in the LOX pilot plant (Ref. 1) was designed to use beneficiated feed rate of 71.29 kg/hr to produce 24 mt of oxygen per year. The reactor required in the current design will use unbeneficiated feed at a rate of 90 kg/hr to produce 5 mt of LOX per year. Since the feed quantity is nearly the same, the overall reactor dimensions of the pilot plant were maintained in the current design. However, several specific details have been modified as required to conform to the needs of the current design goals. The interior diameter of the reactor is 0.31 m at middle bed section. In consideration of the temperature variation existing in the middle (1000°C), bottom (809°C), and top (744°C) beds, the diameter of the reactor is designed to be 0.279 m and 0.267 m at the bottom and top beds, respectively, to maintain the gas velocity at 1 ft/sec in all the three beds. Reactor input material is restricted to sizes greater than 0.045 mm to avoid excessive carryover of fines to avoid clogging up of the system and less than 0.5 mm to allow fluidization to occur. Thus, the residence time of solids is 50 minutes in the top bed, 1.4 hours in the middle bed, and 50 minutes in the bottom bed. The total residence time of the solids in the reactor is 3.08 hrs. The central 0.31 m core of the reactor was surrounded by 7.5 cm of high-density (S.G. 2.24) super-duty firebrick to withstand the erosional nature of the high temperature gas/particles in the fluidized beds. Surrounding this is 23 cm of low-density (S.G. 0.14) insulation. The height of each bed in the reactor has been determined so that there will be enough space above the volume occupied by the solids for disengaging entrained solids. The outside diameter and height of the reactor are 0.9 m and 6.6 m respectively.

Electric Heater

The electric heater heats the gas stream entering the reactor middle bed.

Electrolysis Cell

The solid-state electrolysis cell, operating at 1000°C, separates the hydrogen and water vapor coming out of the reactor into hydrogen and oxygen. The hot hydrogen is recycled to the top bed of the reactor.

Recycle Blower

The recycle blower circulates the hydrogen gas. Gas from the blower enters the reactor at the bottom bed plenum chamber.

Cyclone Separators

The cyclone separators separate the solids from the gases. The cyclone separators contain no moving parts. Dust in the exit gas coming out of each of the reactor beds is removed to prevent the clogging of the system. This dust is returned to the reactor below the surface of each bed.

Solids Settlers

Two solids settlers have been proposed to allow continuous flow of "spent" solids from the bottom bed. Each hopper is sized to allow 2 days residence time for the solids

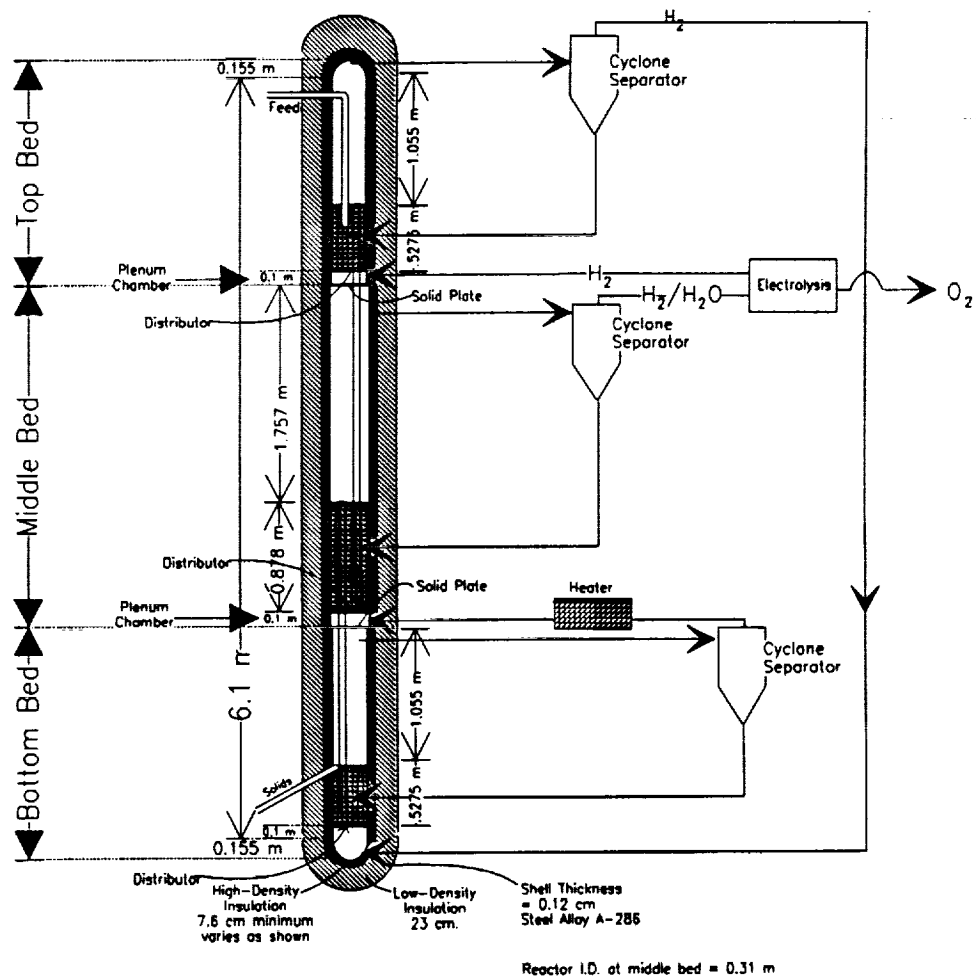


Figure 3
Reactor Details

to settle and separate trapped gases. A gas loop recovers the hydrogen gas and recycles it through the H_2 make-up system. Solids will be allowed into one hopper while the other would be reduced to 0.1 atm. or less to recover gases. This arrangement may also facilitate the addition of a heat recovery system in the future. By putting compressors in series, as shown in Figure 2, the power of each is minimized and redundancy is provided.

Radiator/Thermal Control System (TCS) (not shown)

The radiator/thermal control system rejects the waste heat from oxygen liquefaction equipment. The TCS uses heat exchangers and an appropriate cooling medium. The radiator is positioned to keep the radiator permanently shaded from the sun.

Oxygen Liquefaction and LOX Storage

The oxygen produced by the electrolysis cell is liquefied by a stirling cycle refrigerator. The liquid oxygen is stored in two buried tanks. The re-use of tanks from an abandoned lunar excursion vehicle (LEV) is recommended. This will reduce mass and is a good substitute for new LOX storage tanks brought from Earth.

Hydrogen Make-up System

The hydrogen make-up system includes a tank, a heater, and a compressor. The tank contains enough hydrogen for six months of process leaks and losses. Resupply will be necessary at that time.

Discard Bin

This bin holds 15 days quantity of spent solids for disposal.

Support Structure (not shown)

The process support structure is mounted in the vertical direction to allow gravity flow of the solids in the reactor.

RESULTS

A list of mass, power and dimensions for the equipment required in a 5 mt per year LOX production plant is given in Table 1. The process plant mass, not including mining equipment (front-end loader and hauler) is 5600 kg. The electric power required is 49.0 kW. Plant equipment was sized using the equations, methods and scaling indicators described in Ref. 1.

SUMMARY AND CONCLUSIONS

1. While there are several ways to produce oxygen from lunar materials, the process of reducing ilmenite with hydrogen was considered in this design.
2. Using bulk lunar soil as feedstock without beneficiation components reduces total equipment mass significantly. The mass of the beneficiation components for the 24 mt pilot plant was 3879 kg (Ref. 1). The current design, without beneficiation components, does not require a large-size reactor and the process plant is a viable option. This trade needs to be evaluated for larger production rates.
3. Delivery of small, modular oxygen-production units is easy compared to delivery of a large production plant. Several of the modular units can be deployed and

TABLE 1.
LUNAR OXYGEN PRODUCTION PLANT EQUIPMENT

5 mt/yr, H₂ Reduction of Ilmenite
Ilmenite in regolith = 7.5% wt. No beneficiation.

Equipment	Total Mass (kg)	Total Power (kW)	Dimensions			
			W/D (m)	L (m)	H (m)	V (m ³)
<u>Mining</u>	<i>Rate 1620 mt/yr</i>					
Front-end Loader (1)	2000	0.5	2.1	3.3	2.3	16
Hauler (1)	1000	0.2	2.5	4.0	2.5	25
Mining Total	3000	0.6				
<u>Feed Preparation</u>	<i>Rate 206 kg/hr</i>					
Feed Bin	176	0	4.5	--	3.0	40
Coarse Screen	15	0.5	0.5	0.8	0.4	0.2
Fine Screen	190	6	2.5	3.0	0.5	3.8
<u>Process</u>	<i>Reactor Feed Rate 90 kg/hr</i>					
Pressure Hopper 1	140	--	1.5	--	2.0	3.5
Pressure Hopper 2	140	--	1.5	--	2.0	3.5
Reactor	1960	--	0.9	--	6.6	4.4
Electric Heater	140	25	0.9	1.1	--	0.7
Electrolysis Cell	23	4	0.4	0.4	0.7	0.11
Recycle Blower	29	0.3	0.2	0.2	0.3	0.01
Cyclone Separators (3)	3	--	0.1	--	0.4	0.01
Solids Settlers (2)	288	--	1.2	--	2.0	2.26
Discard Bin	320	--	3.0	--	3.0	21
Oxygen Liquefier	40	1	0.3	0.8	--	0.06
Oxygen Storage Tanks (2)	120	--	1.0	--	--	0.5
Radiator/TCS	400	--	3.0	6.5	0.01	0.2
<u>Hydrogen System</u>						
Liquid Hydrogen Tank	1	0.5	0.5	--	--	0.05
H ₂ Heater	1	0.1	0.1	0.1		--
H ₂ Blower	2	0.1	0.1	--	0.1	--
<u>Process Piping</u>						
In reactor, 4 cm pipe	210	--	--	50	--	--
Others, 0.25 cm pipe	100	--	--	80	--	--
Process Total	4298	37.5				
<u>Margin 30%</u>	1302	11.5				
(includes support structure)						
<u>PLANT TOTALS</u>	5600	49.0				

erected on the Moon to increase the total oxygen production in stages. This may not result in economies of scale, but there are benefits in commonality, operations and construction.

4. A single "point design" to produce 5 mt of LOX per year using unbeneficiated lunar soil shows feasibility and indicates that the reduction of total plant mass is not trivial comparing the 24 mt/yr pilot plant.

5. The 5 mt/yr plant was designed for the feed rate of 90 kg/hr with 7.5% wt ilmenite concentration. If the soil contains a higher percentage of ilmenite (as much as 15% wt.), the process plant will be able to operate without any problem producing more oxygen with the same rate of feed. However, modifications relating to quantity of soil mining, feed rate, heat and power would make the plant operation efficient and effective.

6. Payback of plant mass occurs in one year of operation.

RECOMMENDATIONS

1. The proposed use of a vibrating screen to remove fines of sizes less than 45 microns results in relatively large screen areas, power and mass. An alternate design may use air classifiers which lift the fines and allow the collection of useful material as a "heavy stock". The air classifier with aspirator, cyclone separator, fan and motor may replace the vibrating fine screen. This system can be introduced between top and bottom reactor feed hoppers. A substantial reduction in mass is expected. A detailed pressure analysis and design to incorporate air classifier to the LOX plant is needed. This will also provide the flexibility to change the minimum size limit of the feedstock as required, without replacing any parts. This could be useful if problems occur during operation, or if off-nominal operation is required, as in during start-up and shut-down.

2. Hands-on research to determine the optimum reactor configuration for maximum ilmenite conversion, especially for soil with high ilmenite concentration, is recommended.

3. More detailed design of dust removal and hydrogen removal from the process stream (water and hydrogen) to the electrolysis cell is required to avoid clogging and to reduce the load.

4. At least 90% of the feed to the reactor is rejected as waste at 809°C. Using this waste product and providing only a little more heat, the material can be sintered or even melted and cast into bricks and other products. Radiation and micrometeoroid shield design for the lunar outpost will be economical using cast basalt since the density could be as high as 2900 kg/m³. Casting the co-products of a lunar oxygen plant to form useable materials would be advantageous and may allow their use as radiation protection. This must be explored. Alternately, recovery of this heat value could lower the power requirements of the plant.

REFERENCES

1. Christiansen, Eric L., et al. "Conceptual Design of a Lunar Oxygen Pilot Plant" NAS9-17878. Eagle Engineering, Inc. EEI Report No. 88-182. Houston, Texas. July 1, 1988.
2. Christiansen, Eric L., et al. "Support to Extraterrestrial Propellant Leveraging Trade Study" Eagle Engineering, Inc. EEI Report No. 89-251. Houston, Texas. August 15, 1989.
3. Gibson, M.A. and Knudsen, C.W. " Lunar Oxygen Production from Ilmenite", Lunar Bases and Space Activities of the Twenty-First Century. W.W.Mendell, ed. Lunar and Planetary Institute. 1985
4. Anderson, C., et al. "Conceptual Design of a Technology Verification Unit for Lunar Regolith Volatiles Extraction", Senior Project Final Report, Rice University, April 20, 1990.

57-33
N91-27110-22
P-15

DESIGN OF OPTIMAL CORRELATION FILTERS FOR HYBRID VISION SYSTEMS

Final Report

NASA/ASEE Summer Faculty Fellowship Program -- 1990

Johnson Space Center

Prepared By:

Periasamy K. Rajan, Ph. D.

Academic Rank:

Professor

University & Department:

Tennessee Technological University
Department of Electrical Engineering
Cookeville, TN 38505

NASA/JSC

Directorate:

Engineering

Division:

Tracking and Communications

Branch:

Tracking Techniques

JSC Colleague:

Richard D. Juday, Ph. D.

Date Submitted:

August 15, 1990

Contract Number:

NGT-44-005-803

ABSTRACT

Research is underway at the NASA Johnson Space Center on the development of vision systems that recognize objects and estimate their position by processing their images. This is a crucial task in many space applications such as autonomous landing on Mars sites, satellite inspection and repair, and docking of space shuttle and space station. Currently available algorithms and hardware are too slow to be suitable for these tasks. Electronic digital hardware exhibits superior performance in computing and control; however, they take too much time to carry out important signal processing operations such as Fourier transformation of image data and calculation of correlation between two images. Fortunately, because of the inherent parallelism, optical devices can carry out these operations very fast, though they are not quite suitable for computation and control type of operations. Hence, investigations are currently being conducted on the development of hybrid vision systems that utilize both optical techniques and digital processing jointly to carry out the object recognition tasks in real time.

The author of this report, during his tenure as a summer faculty fellow at the Johnson Space Center studied the various aspects of this research. He collaborated with Dr. Richard juday and his colleagues on developing algorithms for the design of optimal filters for use in hybrid vision systems. Specifically, an algorithm was developed for the design of real-valued frequency plane correlation filters. Further, research was also conducted on designing correlation filters optimal in the sense of providing maximum signal-to-noise ratio when noise is present in the detectors in the correlation plane. Algorithms were developed for the design of different types of optimal filters: complex filters, real-valued filters, phase-only filters, ternary-valued filters, coupled filters. This report presents some of these algorithms in detail along with their derivations.

INTRODUCTION

Recognition of objects by machines is a vital task in many applications related to NASA missions. Docking of space shuttle and space station, extra-vehicular activities by robots, and autonomous landing on Mars are some examples of such applications. To carry out the tasks efficiently and accurately in real time, suitable technologies are being investigated. Most of these tasks involve some form of processing of images acquired using a camera. A straightforward approach is to use electronic digital hardware to carry out the necessary processing involved in the recognition of objects. Researchers in many universities and research laboratories around the world are working on the development and evaluation of algorithms for such processing and pattern recognition. One major drawback of the electronic processing technique is that at the current technology, the algorithms take too much time to be suitable for real time applications.

An alternate technology that is under investigation by many researchers is the use of optical processing. Since through optical techniques processing can be done in a parallel mode, very high speed can be achieved. Taking Fourier transformation, multiplication of a signal by a filter, and obtaining correlation function of two images are some of the operations that optical processing can efficiently perform. However, some operations such as storage and retrieval of data, arithmetic operations, etc., cannot so easily be done using optical techniques. Hence, a hybrid system employing optical techniques for correlation filtering and electronic digital techniques for image storage, retrieval, correlation plane processing, etc., is being investigated for development and use in vision systems.

One major area of investigation in the development of hybrid vision systems is the design of optimal filters that yield high signal to noise ratio (SNR) facilitating accurate detection of objects in the presence of noise and other extraneous objects. During the summer fellowship period, the author collaborated with his JSC colleague Dr. Richard Juday and Dr. B.V.K. Vijaya Kumar of Carnegie Mellon University and developed algorithms for the design of several classes of optimal filters: complex filters, real-valued filters, ternary valued filters, phase only filters, and coupled filters. Some of these algorithms and their derivations will be presented in the following sections.

CORRELATION FILTERS

The theory of correlation filters for the recognition of known objects has been well discussed in the literature [1]. It basically involves the following steps. From a knowledge of the reference image $s(x)$ (we will use 1-D notation for the sake of simplicity), a filter function $h(x)$ is designed. When an image $t(x)$ is to be tested to find whether the reference image is present in the test image or not, the image $t(x)$ is convolved with the filter $h(x)$ to yield $c(x)$. If $c(x)$ has a well pronounced peak above a preassigned threshold, then it is concluded that the reference image is present in the test image; otherwise it is not. If a peak is present, from the position of the peak, one may also estimate the location of the reference image in the test image. When $h(x)$ is equal to $s(-x)$, the matched filter of the reference image, it can be shown that $c(x)$ will yield the cross correlation between the test image and the reference image. Hence, these filters are called correlation filters.

The implementation of correlation filters in the space domain is quite time consuming. An alternate approach is to transform the test and reference images into the Fourier frequency domain, multiply the transforms and then perform the inverse transform to obtain the filter output $c(x)$. If one uses digital techniques using electronic hardware, the Fourier transformation is also equally time consuming. However, Fourier transformation can easily and almost instantaneously be carried out using optical techniques. As a result, considerable research is being conducted on the design of optical correlators for real time pattern recognition purposes. Since the invention of the holographic matched filter for optical correlations by VanderLugt [2] in 1964, many other filters with improved performance have been proposed. The accuracy of detection of these correlation filters depends to a large extent on the absence of noise in an image. It may be noted that the nonreference image

components present in an image may also be considered as noise. To account for the presence of noise and to measure the performance of a filter in the presence of noise, signal-to-noise ratio (SNR) has been introduced and procedures have been devised to design optimal filters that maximize the SNR and thus improve the accuracy of identification.

Derivation of maximum SNR filters of different types (complex, phase only, ternary and coupled filters) in the presence of image-plane noise but no detector noise has been presented in literature. For the case of real filters, derivation of optimal filters starting from the fundamental expression for SNR has not been done so far. We first present such a derivation obtained during the fellowship period. This result has been submitted for publication [3].

OPTIMAL REAL CORRELATION FILTERS

Let $s(x)$ denote the reference image and $S(f)$ its Fourier transform. Let $H(f)$ denote the filter transform. In the absence of input noise the resulting correlator output at the origin is given by

$$c(0) = \int S(f)H(f)df \quad (1)$$

where the limits of integration are those implied by the bandwidths of $S(f)$ and $H(f)$ (whichever has a smaller bandwidth.) A model for possible uncertainties in the input is the additive noise $n(x)$. We model this as a sample realization with mean μ_n and power spectral density $P_n(f)$. The additive noise $n(x)$ in the input leads to randomness in the output $c(0)$. We can show that

$$E\{c(0)\} = \mu_n H(0) + \int S(f)H(f)df \quad (2)$$

and

$$Var\{c(0)\} = \int P_n(f)|H(f)|^2 df \quad (3)$$

When the input to the correlator is only noise $n(x)$, the output $c(0)$ will be a random variable with mean $\mu_n H(0)$ and variance as given in Eqn. (3). For good detection, we need to separate the two means as much as possible while keeping the variance small. A convenient measure for this is the signal-to-noise ratio (SNR) defined as below:

$$SNR = \frac{|\int S(f)H(f)df|^2}{\int P_n(f)|H(f)|^2 df} \quad (4)$$

We will next derive the optimum real filter $H(\cdot)$ that maximizes the SNR. The derivation can be done in more than one way. In the following we use discrete arguments and partial differential equations. For the continuous argument case one may employ variational calculus and obtain identical results. One may also employ Cauchy-Schwartz inequality to obtain the same result. We use such a technique in the next section where detector noise is taken into account.

Let us sample the frequency domain quantities at intervals Δf to produce amplitudes A_k and phases ϕ_k :

$$\begin{aligned} A_k \exp(j\phi_k) &= S(k \Delta f) \\ H_k &= H(k \Delta f) \\ P_{nk} &= P_n(k \Delta f) \end{aligned} \quad (5)$$

Then

$$\text{SNR} = \frac{\left[\sum_{\ell} A_{\ell} H_{\ell} \exp(+j\phi_{\ell}) \right] \left[\sum_k A_k H_k \exp(-j\phi_k) \right]}{\sum_k H_k^2 P_{nk}} \quad (6)$$

where ℓ and k are summed over values appropriate to the filter bandwidth. For an extremum SNR, each filter value H_m is chosen so that the SNR is stationary; i.e., subject to the constraint

$$\frac{\partial}{\partial H_m} (\text{SNR}) = 0 \quad (7)$$

Defining β and B real ≥ 0 such that

$$B \exp(j\beta) = \sum_k A_k H_k \exp(j\phi_k) \quad (8)$$

and taking the partial derivative in (7) we get

$$H_m P_{nm} B^2 = A_m \left[\sum_k H_k^2 P_{nk} \right] B \cos(\phi_m - \beta) \quad (9)$$

We have, then,

$$H_m \propto \frac{A_m}{P_{nm}} \cos(\phi_m - \beta) \quad (10)$$

with proportionality constant independent of m . The SNR is independent of a constant multiplier of H_k , so we may as well make Eqn. (10) an equality. Then we need only solve for β . Eqn. (10) indicates several interesting things. As might have been expected, the optimum-SNR real filter value is directly proportional to the local amplitude of the reference signal transform and inversely proportional to the local noise power. The novelty in this result is that the filter is also weighted according to how well the signal's local phase (ϕ_m) lines up with the phase (β) of the filtered result.

Now, to determine the filter completely, we need to find the value of β . For this we need to solve the nonlinear equations

$$H_m = \frac{A_m}{P_{nm}} \cos(\phi_m - \beta) \quad (11)$$

where

$$\beta = \text{Arg} \left\{ \sum_k H_k A_k \exp(j\phi_k) \right\}$$

A traditional method of solving these equations is as follows: Choose a trial value for $\beta = \beta_1$. Solve for H_m and see if the result is consistent. That is, if the β calculated using the second expression is equal to β_1 , then we have a stationary SNR. This is similar to the previous [4] approaches where calculating a filter giving an extremum in correlation intensity involved a search over a single parameter. Relating to the present case, we would require a search over β . As a significant advance, we have been able to separate β in Eqn. (11). The search for β is over for real filters. Substituting for H_m in the expression for β in Eqn. (11) we have

$$\begin{aligned} \beta &= \text{Arg} \left\{ \sum_k \frac{A_k}{P_{nk}} \cos(\phi_k - \beta) A_k \exp(j\phi_k) \right\} \\ &= \text{Arg} \left\{ \sum_k \frac{A_k}{P_{nk}} [\exp(j\phi_k) \exp(-j\beta) + \exp(-j\phi_k) \exp(j\beta)] A_k \exp(j\phi_k) \right\} \\ &= \text{Arg} \left\{ \exp(j\beta) \left[\sum_k \frac{A_k^2}{P_{nk}} \exp(j2\phi_k) \right] \exp(-j2\beta) + \sum_k \frac{A_k^2}{P_{nk}} \right\} \end{aligned} \quad (12)$$

The above will be true if and only if

$$\beta = \left(\frac{1}{2} \right) \text{Arg} \left\{ \sum_k \frac{(A_k \exp(j\phi_k))^2}{P_{nk}} \right\} + n \frac{\pi}{2}; \text{ where } n \text{ is an integer} \quad (13)$$

By substituting for β in the expression for SNR, we can show that n should be equal to zero for maximum SNR.

We now give an abbreviated discussion of the derivation of the optimal real filter in the continuous domain. Define u and v as follows.

$$\begin{aligned} u &= \left\{ \int H(f) A(f) \exp[+j\phi(f)] df \right\} \left\{ \int H(f) A(f) \exp[-j\phi(f)] df \right\} \\ v &= \int H^2(f) P_n(f) df \\ \text{SNR} &= \frac{u}{v} \end{aligned} \quad (14)$$

(The region of integration, ordinarily a single diffraction order, is implicit.) Following the variational calculus approach given in more detail elsewhere [4], we denote the arbitrary disturbance function as $\mu(\cdot)$, so that the deviated value of the optimum real filtering function is $H(\cdot) + \alpha\mu(\cdot)$ for a (presumed small) scalar, α . The deviated value of the correlation field is obtained by replacing $H(\cdot)$ by $H(\cdot) + \alpha\mu(\cdot)$. Let u_α be u evaluated with $H(\cdot) + \alpha\mu(\cdot)$ replacing $H(\cdot)$ and similarly for v_α . The first-order variation in SNR induced by $\alpha\mu(\cdot)$ is

$$\delta(\text{SNR}) = \alpha \left[\frac{\partial}{\partial \alpha} \left(\frac{u_\alpha}{v_\alpha} \right) \right]_{\alpha=0} . \quad (15)$$

That a given function $H(\cdot)$ produces an extremum of SNR is stated as

$$\delta(\text{SNR}) = 0, \text{ implying } v \delta u - u \delta v = 0 . \quad (16)$$

A relatively straightforward extension of the method shown earlier leads to a form very similar to the discrete case.

$$H(f) = \frac{A(f)}{P_n(f)} \cos[\phi(f) - \beta_1] \quad (17)$$

$$\beta_2 = \text{Arg} \int H(f) A(f) \exp[j\phi(f)] df$$

We have consistency if $\beta_1 = \beta_2$. These expressions are of the same form as those obtained for the discrete case with the replacement of integration in place of summation. Then applying the same development as in Eqns. (12) - (13), we get

$$\beta = \left(\frac{1}{2} \right) \text{Arg} \left\{ \int \frac{[S(f)]^2}{P_n(f)} df \right\} \quad (18)$$

$$= \left(\frac{1}{2} \right) \text{Arg} \left\{ \int \frac{[S_R(f)]^2 - [S_I(f)]^2}{P_n(f)} df + j2 \int \frac{S_R(f) S_I(f)}{P_n(f)} df \right\}$$

We will now consider some special cases. The discussion in the following is restricted to the continuous case. Results for the discrete case are exactly analogous, as is apparent from the congruence between Eqns. (11) and (17).

$$\text{i) } s(x) \text{ is real. Then } \int \frac{S_R(f) S_I(f)}{P_n(f)} df = 0 ; \quad \therefore \beta = 0 \text{ or } \frac{\pi}{2}. \quad (19)$$

$$\text{ii) } s(x) \text{ is real and } \int \frac{[S_R(f)]^2}{P_n(f)} df = \int \frac{[S_I(f)]^2}{P_n(f)} df. \text{ Then } \beta \text{ is arbitrary.} \quad (20)$$

$$\text{iii) } s(x) \text{ is complex and } \int \frac{[S_R(f)]^2}{P_n(f)} df = \int \frac{[S_I(f)]^2}{P_n(f)} df. \quad \text{Then } \beta = \frac{\pi}{4} \quad (21)$$

Cases i) and ii) correspond to the cases considered by Kumar [5] and case iii) is a new result.

The SNR of the real filter, obtained by substituting the optimal value of $H(f)$ and simplifying the expression, is given by

$$SNR_{RMF} = \frac{1}{2} \int \frac{|S(f)|^2}{P_n(f)} df + \frac{1}{2} \left| \int \frac{[S(f)]^2}{P_n(f)} df \right| \quad (22)$$

where SNR_{RMF} represents the SNR of the real matched filter. A relative measure of performance of the real filter in comparison with the complex matched filter is obtained as

$$\frac{SNR_{RMF}}{SNR_{CMF}} = \frac{1}{2} + \frac{1}{2} \frac{\left| \int \frac{[S(f)]^2}{P_n(f)} df \right|}{\int \frac{|S(f)|^2}{P_n(f)} df} \quad (23)$$

where SNR_{CMF} represents the SNR of the complex matched filter. In the above expression we see that the SNR of the optimal real filter is less than or equal to the optimal complex filter. However, the maximum loss that results from the use of our optimal real filter is 3 dB.

Figure 1 in the earlier paper [5] shows the loss in SNR (relative to the matched filter) when using the optimal real filter. That is valid here also. Only what is meant by "normalized even-part energy" changes depending on $P_n(f)$. As shown before, we lose at most 3 dB in SNR (compared with the matched filter) when we use the optimum real filter. This discussion also shows that the optimal real filter is not the amplitude part of the complex matched filter.

DETECTOR NOISE AND SNR

When the correlation output $c(0)$ is detected by a photodetector, several things happen. Detectors respond only to $|c(0)|$, thus ignoring all phase information. Also, the detectors introduce a gain and some noise. An accurate model for detector noise is complicated and must include the signal dependent nature of detector noise. Instead, we will use the following simple model for y , the detector output:

$$y = c(0) + n_d \quad (24)$$

In this detector noise model, we assume (without loss of generality) that the detector gain is unity and that the detector noise n_d is additive. The noise n_d is assumed to have a mean μ_d and variance σ_d^2 . The additive assumption is somewhat questionable. However, it makes the analysis tractable and it helps us to illustrate the main point we want to emphasize (i.e., we must trade-off input noise tolerance for detector noise tolerance). We will now find the mean and variance of y in Eqn. (24) for the two possible input cases.

When the input contains only noise $n(x)$ let the output be y_0 . Then the mean and variance of y_0 are given as follows:

$$E\{y_0\} = \mu_d + \mu_n \quad (25)$$

and

$$Var\{y_0\} = \sigma_d^2 + \int P_n(f) |H(f)|^2 df \quad (26)$$

When the input contains signal $s(x)$ corrupted by additive noise $n(x)$, let the output be y_1 . Then the mean and variance of y_1 are given as below:

$$E\{y_1\} = \mu_d + \mu_n + \int S(f) H(f) df \quad (27)$$

and

$$Var\{y_1\} = \sigma_d^2 + \int P_n(f) |H(f)|^2 df \quad (28)$$

Using the statistics in Eqns. (25)-(28), the SNR in the presence of detector noise can be expressed as

$$\begin{aligned} SNR &= \frac{|E\{y_1\} - E\{y_0\}|^2}{\frac{1}{2} (Var\{y_1\} + Var\{y_0\})} \\ &= \frac{|\int S(f) H(f) df|^2}{\sigma_d^2 + \int P_n(f) |H(f)|^2 df} \end{aligned} \quad (29)$$

Note that the only difference between the SNR expressions in Eqn. (4) and Eqn. (29) is the extra σ_d^2 in the denominator of Eqn. (29). However this makes the optimal filter choices for the two SNRs different. When σ_d^2 is very small (compared to the input noise term), the two SNRs are identical and the previous optimal filters will still be optimal. However, when σ_d^2 is very large (such that the input noise term can be ignored) SNR is simply proportional to $|E\{c(0)\}|^2$ and we must simply maximize the correlation value at the center. In the next section we derive expressions for $H(f)$ that maximize the SNR in Eqn. (29).

OPTIMAL FILTERS WHEN DETECTOR NOISE IS PRESENT

In this section we will determine $H(f)$ that maximize the SNR in Eqn. (29). There are five different cases: (1) Complex filters, (2) real filters, (3) phase only filters (4) binary phase only filters and (5) coupled filters. In each case while determining these filters, we must use the condition

$$|H(f)| \leq 1 \quad (30)$$

since the transmittance of an optical filter can never exceed unity. Also, without this constraint, the detector noise variance σ_d^2 will not be meaningful.

Complex Filters

Let us allow $H(f)$ to be complex. Let the filter energy E_h be defined as follows:

$$E_h = \int |H(f)|^2 df \quad (31)$$

Then the SNR in Eqn. (29) can be written as below:

$$SNR = \frac{|\int S(f)H(f) df|^2}{\int [\sigma_{id}^2 + P_n(f)] |H(f)|^2 df} \quad (32)$$

where

$$\sigma_{id}^2 = \frac{\sigma_d^2}{E_h} \quad (33)$$

is the variance of an equivalent white noise at the input which gives the same effect at the origin of the correlation plane as the detector noise with variance σ_d^2 for the given filter $H(f)$.

To find the optimal choice of $H(f)$ we now apply the Cauchy-Schwartz inequality to the numerator to get the following:

$$\begin{aligned} |\int S(f)H(f) df|^2 &= \left| \int \frac{S(f)}{\sqrt{\sigma_{id}^2 + P_n(f)}} \left[H(f) \sqrt{\sigma_{id}^2 + P_n(f)} \right] df \right|^2 \\ &\leq \left(\int \frac{|S(f)|^2}{[\sigma_{id}^2 + P_n(f)]} df \right) \left(\int |H(f)|^2 [\sigma_{id}^2 + P_n(f)] df \right) \end{aligned} \quad (34)$$

Substituting Eqn. (34) in Eqn. (32), we obtain the following result:

$$SNR \leq \int \frac{|S(f)|^2}{\sigma_{id}^2 + P_n(f)} df = SNR_{\max}(\sigma_{id}^2) \quad (35)$$

with equality occurring if and only if

$$H(f) = \alpha \frac{S^*(f)}{\sigma_{id}^2 + P_n(f)} \quad (36)$$

where α is chosen to satisfy the constraint that the maximum magnitude of the filter is unity. Then the optimal filter $H(f)$ can be written as

$$H_{\text{opt}}(f) = \frac{S^*(f)/[\sigma_{\text{id}}^2 + P_n(f)]}{|S^*(f)/[\sigma_{\text{id}}^2 + P_n(f)]|_{\text{max}}} \quad (37)$$

As σ_{id}^2 is related to the detector noise σ_d^2 , this filter will also be optimal for a detector noise $\sigma_d^2 = \sigma_{\text{id}}^2 E_h$ where

$$E_h = \int |H(f)|^2 df \quad (38)$$

Thus starting with σ_{id}^2 , we can design an optimal filter and find the corresponding detector noise for which this filter is the optimal one. To design a filter for a given σ_d^2 , we may use either an optimization scheme or do the following: Obtain a graph relating σ_{id}^2 and σ_d^2 by designing a number of filters with different values of σ_{id}^2 . Then from the graph, a suitable σ_{id}^2 is picked for a given σ_d^2 . Once we know σ_{id}^2 , we can design the optimal filter. It may be verified that for the special case of $P_n(f) = \text{a constant}$, the classical matched filter is still the optimal filter.

Real Filters

Let us now consider the case when $H(f)$ is real. Then using the same notation for E_h as above, we have

$$SNR = \frac{|\int S(f)H(f)df|^2}{\sigma_d^2 + \int P_n(f)|H(f)|^2 df} = \frac{|\int S(f)H(f)df|^2}{\int (\frac{\sigma_d^2}{E_h} + P_n(f))|H(f)|^2 df} = \frac{cc^*}{\int P_{\text{nt}}(f)|H(f)|^2 df} \quad (39)$$

where

$$c = \int S(f)H(f)df, \text{ and } P_{\text{nt}} = \frac{\sigma_d^2}{E_h} + P_n. \quad (40)$$

Expressing cc^* as

$$cc^* = \left(\frac{cc^* + c^*c}{2|c|} \right)^2 \quad (41)$$

we get

$$SNR = \frac{\left(\int \frac{cS^*(f)H^*(f) + c^*S(f)H(f)}{2|c|} df \right)^2}{\int P_{\text{nt}}(f)|H(f)|^2 df} \quad (42)$$

As for real filters $H^* = H$, we can write the expression for SNR as

$$SNR = \frac{\left(\int \frac{(cS^*(f) + c^*S(f))}{2|c|} H(f) df \right)^2}{\int P_{nt}(f) |H(f)|^2 df} \quad (43)$$

Applying Cauchy-Schwartz inequality we get

$$SNR = \frac{\left(\int \frac{cS^*(f) + c^*S(f)}{2|c|\sqrt{P_{nt}(f)}} \sqrt{P_{nt}(f)} H(f) df \right)^2}{\int P_{nt}(f) |H(f)|^2 df} \leq \int \left(\frac{cS^*(f) + c^*S(f)}{2|c|\sqrt{P_{nt}(f)}} \right)^2 df \quad (44)$$

Equality (and hence maximum SNR) is achieved iff

$$H(f) = \alpha \frac{cS^*(f) + c^*S(f)}{2|c|\sqrt{P_{nt}(f)}} = \alpha \frac{|S(f)|}{\sqrt{P_{nt}(f)}} \cos(\phi_s(f) - \beta) \quad (45)$$

where $\phi_s(f) = \text{Arg}(S(f))$, $\beta = \text{Arg}(c)$ and α is a real constant such that the maximum response condition is satisfied. To determine β , substitute the optimal $H(f)$ in the expression for c (Eqn. (40)) and simplify as

$$|c|e^{j\beta} = \frac{\alpha}{2} \left(\int \frac{|S(f)|^2}{P_{nt}(f)} df + e^{-j2\beta} \int \frac{[S(f)]^2}{P_{nt}(f)} df \right) e^{j\beta} \quad (46)$$

The above condition will be satisfied if and only if

$$\beta = \frac{1}{2} \left(\text{Arg} \int \frac{[S(f)]^2}{P_{nt}(f)} df + n\pi \right) \quad (47)$$

where n is an integer. Further we can show that SNR will be larger if $n = 0$ (or any even integer) than when n is odd. Thus β value is given by

$$\beta = \frac{1}{2} \text{Arg} \left(\int \frac{[S(f)]^2}{P_{nt}(f)} df \right) \quad (48)$$

With this β value the maximum SNR value can be shown to be

$$SNR_{\max} = \frac{1}{2} \int \frac{|S(f)|^2}{P_{nt}(f)} df + \frac{1}{2} \left| \int \frac{[S(f)]^2}{P_{nt}(f)} df \right| \quad (49)$$

The optimal $H(f)$ satisfying the maximum condition is given by

$$H_{\text{opt}}(f) = \frac{\frac{|S(f)|}{P_{\text{nt}}(f)} \cos(\phi_s(f) - \beta)}{\left| \frac{|S(f)|}{P_{\text{nt}}(f)} \cos(\phi_s(f) - \beta) \right|_{\text{max}}} \quad (50)$$

The above $H(f)$ is optimum for a given input noise. To determine $H(f)$ for a given detector noise one can follow a procedure similar to that used in the complex filter case of either using an optimization scheme or setting up a graph between σ_{id}^2 and σ_{d}^2 and selecting a suitable σ_{id}^2 for a given σ_{d}^2 .

Phase Only Filter

Let $H(f)$ be a phase-only filter (POF) with a region of support R defined below:

$$H(f) = \begin{cases} e^{j\phi_h(f)} & \text{for } f \in R \\ 0 & \text{for } f \notin R \end{cases} \quad (51)$$

Substituting this in Eqn. (29) we obtain the following expression for SNR:

$$SNR = \frac{\left| \int_R |S(f)| e^{j[\phi_s(f) + \phi_h(f)]} df \right|^2}{\sigma_{\text{d}}^2 + \int_R P_{\text{n}}(f) df} \quad (52)$$

The denominator in Eqn. (52) does not depend on $\phi_h(f)$ and the numerator is maximized by choosing $\phi_h(f) = -\phi_s(f)$, i.e., the conventional POF maximizes the SNR in the presence of detector noise. The next task is to find the optimal region of support R . Let A_R denote the area of the region R . Then the SNR in Eqn. (52) can be rewritten as follows:

$$SNR = \frac{\left(\int_R |S(f)| df \right)^2}{\int_R \left(\frac{\sigma_{\text{d}}^2}{A_R} + P_{\text{n}}(f) \right) df} \quad (53)$$

The SNR expression in Eqn. (53) is identical to the SNR expression derived elsewhere [6] for the case of no detector noise and input colored noise with spectral density $[P_{\text{n}}(f) + \sigma_{\text{d}}^2 / A_R]$. It is shown there that the region of support R that maximizes the SNR must be of the form:

$$R = \left\{ f : \frac{|S(f)|}{\left(\frac{\sigma_{\text{d}}^2}{A_R} + P_{\text{n}}(f) \right)} \geq T \right\} \quad (54)$$

where T is an unknown threshold. Since A_R depends on R , Eqn. (54) is an implicit equation. An algorithmic procedure for determining R is given below:

To illustrate the workings of the algorithm, let us assume that $S(f)$ is represented by an array of $64 \times 64 = 4096$ pixels. Each frequency in this array may or may not be included in R . Thus there are 2^{4096} possible choices of R from which we must select the one that maximizes the SNR. Exhaustive search is obviously impossible, but the characterization in Eqn. (54) proves useful. For example, there are 4096 different choices for R with $A_R = 1$. Among all of them, the best R is the one that includes the highest value of $|S(f)|/[\sigma_d^2 + P_n(f)]$. Extending this to other A_R values, we get the following algorithm:

Step 0: Start with $A_R = 1$.

Step 1: Compute

$$G(f) = \frac{|S(f)|}{\frac{\sigma_d^2}{A_R} + P_n(f)} \quad (55)$$

Step 2: Arrange sampled values of $G(f)$ in the descending order as $G_1 > G_2 > \dots > G_N > 0$.

Step 3: Construct the optimal $R^*(A_R)$ of area A_R by including pixels corresponding to G_1, G_2, \dots, G_{A_R} .

Step 4: Compute the $\text{SNR}(A_R)$ obtained by using $R^*(A_R)$. If A_R is greater than or equal to N , go to Step 5. Otherwise go to step 2.

Step 5: Determine the largest $\text{SNR}(A_R)$. This determines the optimal region of support.

The above algorithm involves N sorts where each sort is of N values. Since A_R changes by only 1 from one sort step to the next, we do not expect that the sorting order will change significantly from one step to the next. This can be used to speed up the algorithm still further.

Ternary-Valued and Coupled Filters

Optimal ternary-valued filters (with values $-1, +1$, and 0) in the presence input and detector noises have also been derived during the fellowship period. However due to lack of space, they are not discussed in this report. Similarly, the design of optimal coupled filters in the presence of input and detector noises have also been investigated. These results will be sent for publication in due course.

SUMMARY

In this report, the derivation of optimal real correlation filters when detector noise is not present is first presented. Then a model for the detector noise is presented. Using this model, the optimal filter design problem when detector noise is present is formulated. Derivations for the optimal filter when the filter is an unconstrained complex filter, real filter, and phase only filter are presented. Algorithms for their design are also presented. Implementation and testing of these algorithms will be taken up during the follow-up grant period.

REFERENCES

- [1] H. Bartlett, Unconventional correlators, Chapter 2.2 in Optical Signal Processing, J.L. Horner (Ed.), Academic Press, Inc., San Diego, 1987.
- [2] A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Transactions on Information Theory, Vol. IT-10, pp. 139-184, 1964.
- [3] Richard D. Juday, B.V.K. Vijaya Kumar and P.K. Rajan, "Optimal real correlation filters," Applied Optics, in review, 1990.
- [4] Richard D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," Applied Optics, Vol. 28, pp. 4865-4869, 1989.
- [5] B.V.K. Vijaya Kumar, "Signal-to-noise ratio loss in correlators using real filters," Applied Optics, Vol. 28, pp. 3287-3288, 1989.
- [6] Zouhir Bahri and B.V.K. Vijaya Kumar, Fast algorithms for designing optimal phase-only filters (POFs) and binary phase-only filters (BPOFs)," Applied Optics, Vol. 29, pp. 2992-2996, July 1990.

58-39
20223
N91-27 P116

DEVELOPMENT OF LOAD-DEPENDENT RITZ VECTOR METHOD
FOR STRUCTURAL DYNAMIC ANALYSIS OF LARGE SPACE STRUCTURES

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1990

Johnson Space Center

CD 345132

Prepared By:	James M. Ricles, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	University of California, San Diego Department of AMES, R-011 La Jolla, California 92093
NASA/JSC	
Directorate:	Engineering
Division:	Structures and Mechanics
Branch:	Loads and Structural Dynamics
JSC Colleague:	David A. Hamilton
Date Submitted:	August 10, 1990
Contract Number:	NGT-44-005-803

ABSTRACT

In the analysis of large space structures, such as the Space Shuttle and Space Station Freedom, analytical models are dealt with which have a large number of degrees of freedom (DOF). It is important that the method used to analyze these structure be as accurate and efficient as possible in order that structural margins and areas requiring redesign can be identified and subsequent analysis be performed. This study involved the development and preliminary assessment of a method for dynamic structural analysis based on load-dependent Ritz vectors. The vector basis is orthogonalized with respect to the mass and structural stiffness in order that the equations of motion can be uncoupled and efficient analysis of large space structure performed. A series of computer programs was developed based on the algorithm for generating the orthogonal load-dependent Ritz vectors. Transient dynamic analysis performed on the Space Station Freedom using the software was found to provide solutions that required a smaller number of vectors than the modal analysis method. Error norms based on the participation of the mass distribution of the structure and spatial distribution of structural loading, respectively, were developed in order to provide an indication of vector truncation. These norms are computed before the transient analysis is performed. An assessment of these norms through a convergence study of the structural response was performed. The results from this assessment indicate that the error norms can provide a means of judging the quality of the vector basis and accuracy of the transient dynamic solution.

INTRODUCTION

The finite element method is the standard method for developing models of space structures for structural analysis. The finite element models of present and future space structures, such as the Space Shuttle and planned Space Station Freedom, possess a large number of degrees of freedom (DOF). To calculate the response of these structures for dynamic loading conditions involves using the finite element models in a transient dynamic or random structural vibration analysis. The standard procedure for doing this utilizes the modal analysis method. For structural models with a large number of DOF the modal vector basis is generally truncated because of the enormously large amount of computation effort and time required to calculate all eigenvectors and eigenvalues of the finite element model. The constraint of having to use a truncated modal basis and the fact that the computational effort to calculate vibration characteristics is costly compared to Ritz based methods [Bathe, 1982] gives motivation and justification for considering other procedures for generating an orthogonal vector basis.

This study is concerned with the development and preliminary assessment of a load-dependent Ritz vector method for use in structural dynamic analysis of large space structures. The scope of work in this study included developing an algorithm which was efficient in terms of numerical operations and has the ability to be used in a production environment. The algorithm was coded on the CRAY X-MP/EA 464 Supercomputer using the FORTRAN computer language. A transient dynamics analysis of the Space Station Freedom was performed using the computer program in order to assess the algorithm. This assessment involved comparing these results to those from a modal analysis. A summary of this assessment is presented in this report. Conclusions based on the finding of the analysis performed are noted and topics for future research are suggested. Also included in this report is a description of the series of computer programs developed which allow one to perform a transient dynamic load-dependent Ritz vector based analysis using NASTRAN [CSA/NASTRAN, 1988] finite models and to post process the results.

LOAD-DEPENDENT RITZ VECTOR ALGORITHM

Block Algorithm and Error Norms

The algorithm for generating an orthogonal load-dependent vector basis

is based on using the static amplitudes of the dynamic loads as selected times. The vector basis generated aligns itself with respect to the loading, consequently the vector basis has the potential of a high participation with respect to the response to the dynamic loading. A summary of the algorithm is given in Fig. 1. The Ritz vectors are generated in blocks, each block having several vectors, in order to deal with closely spaced vibration frequencies and a spatial varying loading case that can often exist in space structures. The algorithm begins with the shifting of the structural stiffness matrix \mathbf{K} using the shift constant σ , followed by the calculation of the set of displacements \mathbf{U}_0 reflecting the response to the block loading $\mathbf{F}(s)$. Shifting of \mathbf{K} is necessary in order to remove the rigid body modes and permit \mathbf{K} to be factorized. Any structural vibration frequencies near the shift point are well represented in the Ritz vector basis. The vectors \mathbf{X} which are generated are orthogonalized with respect to the mass matrix \mathbf{M} and contain a static residual to reduce the effects of truncation of higher frequencies. A final step of orthogonalizing the vectors \mathbf{X} with respect to \mathbf{K} is necessary to produce the vector basis ${}^o\mathbf{X}$ having the frequencies $\hat{\omega}$ in order to uncouple the equations of motion.

The concept of a block form was motivated by its use in previous work by the investigator [Ricles et. al., 1990] and that of other researchers [Nour-Omid and Clough, 1985].

During the generation of vectors \mathbf{X}_i the norm ϵ_{u_i} is computed, and is intended to provide an indication of the participation of the mass distribution. This norm is computed using the displacements \mathbf{U}_i at each cycle and the initial displacement set \mathbf{U}_0 , where at cycle i :

$$\epsilon_{u_i} = \frac{\|\mathbf{M}\mathbf{U}_i\|_{\infty}}{\|\mathbf{M}\mathbf{U}_0\|_{\infty}} \quad (1)$$

For the analysis performed the behavior of ϵ_{u_i} was studied to determine whether it could be used as criteria to judge when to terminate the generation of vectors. A second norm ϵ_L , intended to represent the degree of participation of the spatial load distribution, was also calculated and studied to determine whether it was reliable for use in judging the quality of the vector basis ${}^o\mathbf{X}$ in responding with high participation to the loading. This norm is computed by the following formula, where for k vectors in ${}^o\mathbf{X}$:

$$\epsilon_L = \frac{|\mathbf{P}(t)\mathbf{e}_L|}{\mathbf{P}^T(t)\mathbf{P}(t)} \quad (2)$$

1. Dynamic Equilibrium Equations:

$$\mathbf{M}\ddot{\mathbf{Z}}(t) + \mathbf{C}\dot{\mathbf{Z}}(t) + \mathbf{K}\mathbf{Z}(t) = \mathbf{F}(s)\alpha(t)$$

2. Initial Calculations:

(a) Shift and Factorize Stiffness Matrix

$$\mathbf{K}^* = \mathbf{K} + \sigma\mathbf{M}$$

$$\mathbf{K}^* = \mathbf{L}\mathbf{D}\mathbf{L}^T$$

(b) Solve for Static Response to Block Loading

$$\mathbf{K}^*\mathbf{U}_0 = \mathbf{F}(s)$$

3. Calculate Rigid Body Modes:

(a) Use Geometric Description and DOF Relationship to Describe Rigid Body Motions \mathbf{X}_1

(b) Generate First Block(s) of Orthonormalized Ritz Vectors

$$\mathbf{X}_1 = \beta\bar{\mathbf{X}}_1$$

$$\beta = (\bar{\mathbf{X}}_1^T \mathbf{M} \bar{\mathbf{X}}_1)^{-1}$$

(c) Remove Rigid Body Modes From Static Block \mathbf{U}_0 (Gram-Schmidt Orthogonalization)

$$\mathbf{U}_1 = \mathbf{U}_0 - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{M} \mathbf{U}_0)$$

4. Generate Additional Ritz Vectors $\mathbf{X}_i, i = 2, \dots, n-1$:

(a) Solve for $\bar{\mathbf{X}}_i$

$$\mathbf{K}^*\bar{\mathbf{X}}_i = \mathbf{M}\mathbf{U}_{i-1}$$

(b) M-Orthogonalize $\bar{\mathbf{X}}_i$ against previous blocks (Gram-Schmidt)

$$\mathbf{X}_i^* = \bar{\mathbf{X}}_i - \sum_{j=1}^{i-1} \mathbf{X}_j \mathbf{X}_j^T \mathbf{M} \bar{\mathbf{X}}_i \quad 1 \leq m \leq i-2$$

(c) M-Orthogonalize Vectors in Block \mathbf{X}_i^* by Modified Gram-Schmidt to obtain \mathbf{X}_i

(d) Remove New Ritz Block \mathbf{X}_i From Static Block \mathbf{U}_{i-1} (Gram-Schmidt)

$$\mathbf{U}_i = \mathbf{U}_{i-1} - \mathbf{X}_i (\mathbf{X}_i^T \mathbf{M} \mathbf{U}_{i-1})$$

5. Add Static Block Residual \mathbf{U}_{n-1} as Static Correction Terms \mathbf{X}_n .

6. Make Ritz Vectors \mathbf{X} Stiffness Orthogonal (Optional - Uncouples Equations of Motion):

(a) Form and Solve the Reduced $n \times n$ eigenvalue Problem

$$\mathbf{K}^{**} = \mathbf{X}^T \mathbf{K}^* \mathbf{X}$$

$$[\mathbf{K}^{**} - \omega^2 \mathbf{I}] \Psi = 0$$

(b) Compute Final Ritz Vectors ${}^o\mathbf{X} = \mathbf{X}\Psi$

Fig. 1- Block Form Load-Dependent Ritz Vector Algorithm for Semipositive Definite Systems.

where $\mathbf{P}(t)$ is the load vector at time t and \mathbf{e}_L the error in the representation of the load in Ritz coordinates, with

$$\mathbf{e}_L = \mathbf{P}(t) - \sum_{j=1}^k {}^0\mathbf{X}_j \mathbf{P}^T(t) \mathbf{M} {}^0\mathbf{X}_j \quad (3)$$

Software Development

The computer program named RITZ was developed to generate ${}^0\mathbf{X}$, $\dot{\omega}$, and ϵ_u . RITZ interfaces with the bulk NASTRAN data set and NASTRAN OUTPUT4 [CSA/NASTRAN, 1988] files containing the structural mass and stiffness matrices to define the structural model. To maximize the efficiency, the symmetry and any sparseness of these matrices are exploited in their incore storage and factorization. A second computer program named NAS-DYN was developed which gives the user the choice of calculating either the norm ϵ_L or the transient dynamic response of the structure. Either Ritz vectors or NASTRAN modal vectors can be used. The results of the transient analysis can be post processed by a third newly developed program POSTRN for producing plot files which are compatible with the JSC Structures and Mechanics Division PLOT2D graphics software.

SPACE STATION ANALYSIS

Program of Investigation

The Space Station Freedom was analyzed for a simulated docking with the Space Shuttle. A view of the Station is given in Fig. 2. In the analyses performed the NASTRAN MB15YZ model of the Station with 2803 DOF was used along with docking load case 915L, consisting of the set of transient forces shown in Fig. 3 applied to the end of the docking arm. The position of the PV arrays in this model had an orientation in the Y-Z plane as opposed to the X-Y plane as shown in Fig. 2. RITZ was used to generate vector bases with 30, 60, and 90 vectors and DYNTRN used to compute the elastic displacements and total accelerations of selected DOF based on a constant damping ratio of 0.01. The vector bases were generated using a block size of 6 vectors, with \mathbf{U}_0 consisting of six displacement vectors corresponding to the response of the six individual docking loads acting on the structure at time=1.93 seconds. The response at the tip of an outboard PV array (node 8022 in Fig. 2) is discussed in this report. Neither a complete solution

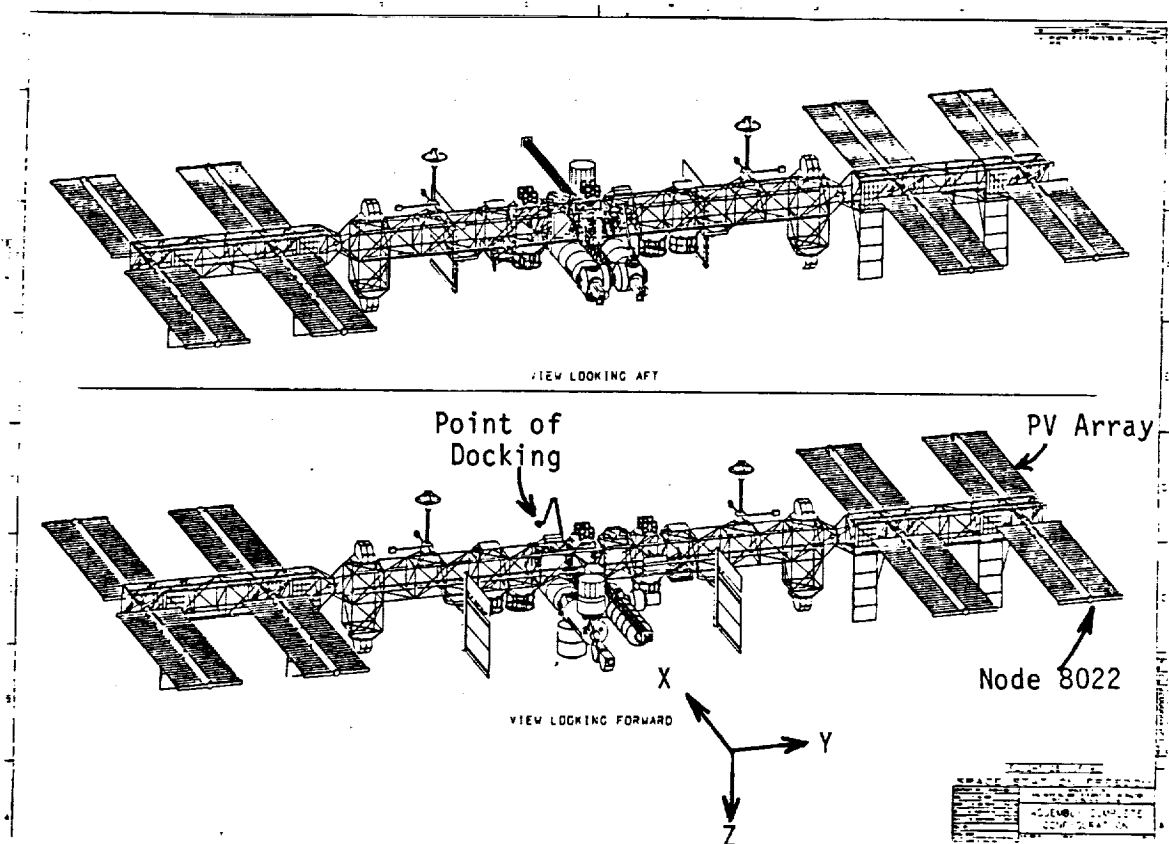


Fig. 2-Space Station Freedom with PV Arrays in The X-Y Plane.

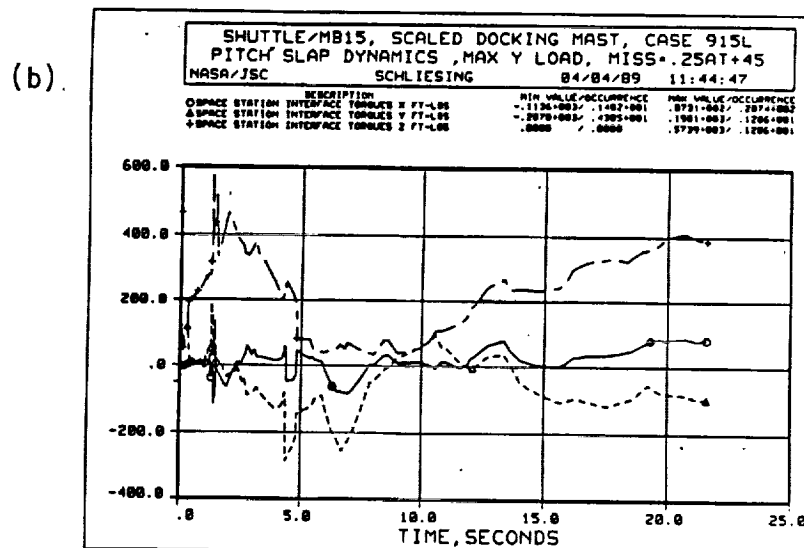
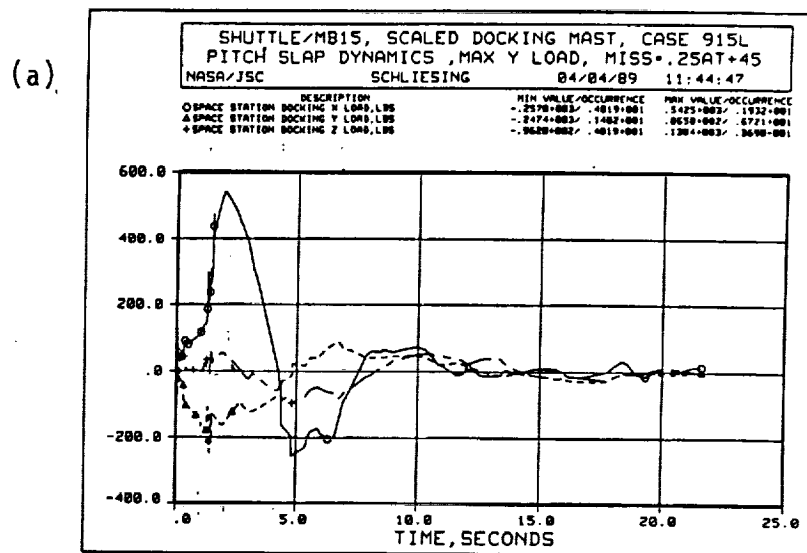


Fig. 3 Docking Load Case 915L (a) Translational, and (b) Rotational Transient Forces.

based on a numerical integration of the coupled equations of motion nor a complete modal vector basis was available to compare with the Ritz solution. For purposes of this preliminary study a comparison was made with solutions obtained from a truncated modal basis consisting of 210 modes. The frequency range of the modes ranged from 0 to 6.0 Hz. Fig. 4 shows a comparison of the frequency range of the modal vectors and selected Ritz vector bases. The 60 Ritz vector basis with a static residual (identified as SR in the legend of Fig. 4) has approximately the same frequency range as the modal basis. It can be seen that a larger Ritz vector basis has a greater span and that this span is extended for those basis containing the static residual as opposed to not having it (identified as NSR in the figure's legend). All Ritz vector basis presented in this report were generated with a shift constant of $\sigma = 1.0$, except for the 60 vector basis with $\sigma = 200.0$ identified in Fig. 4 as SHI. The effect of using a larger value for σ is that the basis has more distantly spaced frequencies and a greater span.

Presentation of Transient Response

A time history plot of the translational displacements and accelerations in the x-direction are shown in Figs. 5 and 6 for the 90 Ritz vector and modal solution. The Ritz based displacements are shown to agree closely with the modal solution, whereas there exists a greater discrepancy between the accelerations computed by the two methods. The maximum discrepancy in the transient translational displacement and acceleration along the X-, Y-, and Z-axes based on the results from the modal vectors and Ritz vectors with a 90, 60, 30 vector basis are shown in Fig. 7. These results correspond to the node at the end of the outboard PV array. It is apparent in Fig. 7 that the discrepancy between the 210 modal vector and Ritz solutions decreases as the number of vectors in the Ritz basis is increased from 30 to 90 vectors. The Ritz solution with a 60 vector basis is able to achieve the same result as the 210 modal vector solution for displacements. The Ritz based solution for accelerations shows greater discrepancies with the modal solution, and requires a larger number of vectors in the basis to achieve the same level as accuracy as that found in the displacements at 60 Ritz vectors.

Analysis of Results

The norm ϵ_u is plotted against the number of blocks of Ritz vectors in

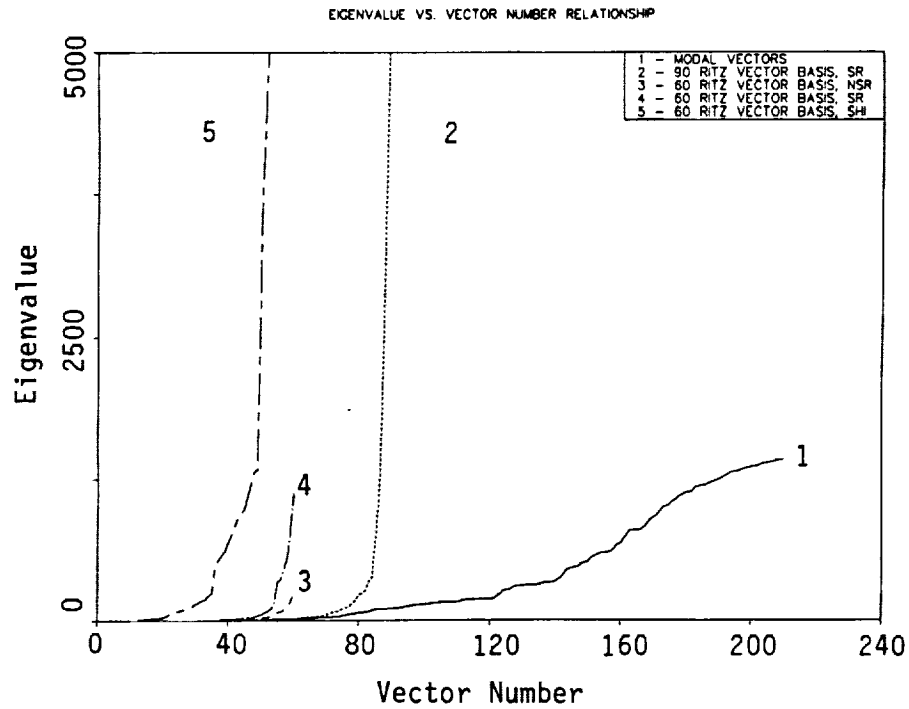


Fig. 4.- Span of Eigenvalue (Frequency) Range of Modal and Ritz Vector Bases.

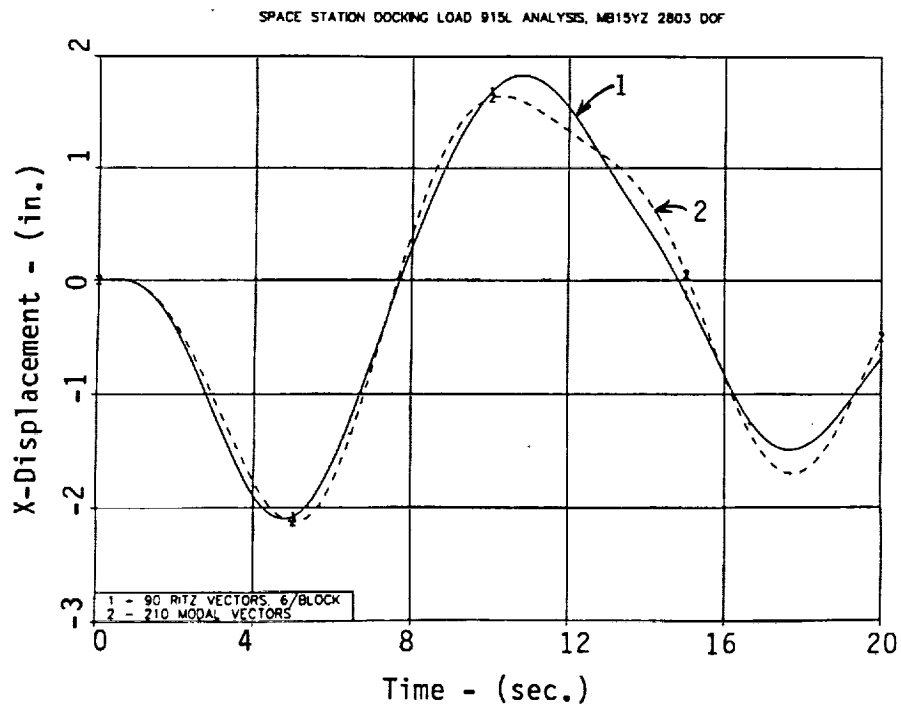


Fig. 5.- Displacement Response at End of Outboard Pv Array.

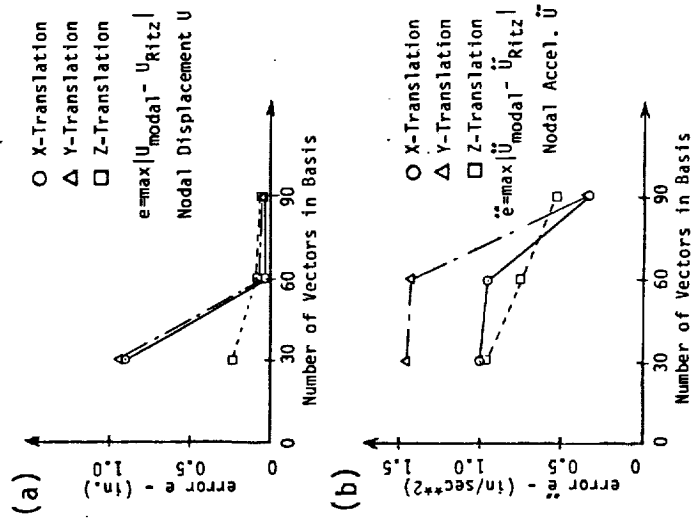


Fig. 7.- Error Between Modal and Ritz solutions for (a) Displacements, and (b) Accelerations at end of PV Array.

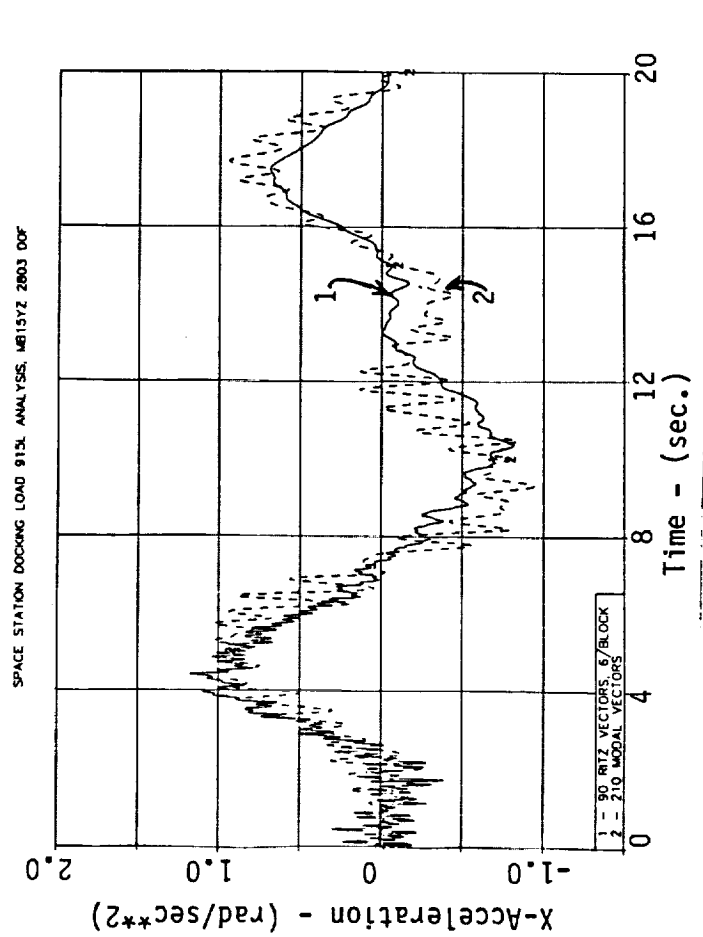


Fig. 6.- Acceleration Response at End of Outboard PV Array.

Fig. 8 to establish whether it can give a good indication of when to terminate the generation of vectors. An examination the data reveals that for 30 Ritz vectors (5 blocks x 6 vectors per block = 30) the value of ϵ_u is approximately 50 times greater than that for 60 Ritz vectors (10 blocks x 6 vectors per block = 60), at which the displacements were in very close agreement with the modal solution. The fact that ϵ_u has become quite small at 60 Ritz vectors suggest that there is not much to be gained by 30 additional vectors to achieve a 90 vector basis. As discussed above this was found to be true only for the displacements. Thus, ϵ_u could be considered more reliable for displacements than accelerations .

The behavior of ϵ_u in Fig. 8 shows that a fluctuation in the curve can exist where a local maximum develops. This phenomenon is quite noticeable for the case of 5 blocks with 6 vectors per block. This behavior is associated the structure responding more to the forces related to the displacements U_i as opposed to U_0 . Numerical experimentation indicated that apparently this occurs when the block loads used to calculate U_0 are not uniformly displacing the mass, leading to displacement patterns U_i in some of the later cycles which 'accelerate' more the mass. This same experimentation indicated that the fluctuation tends to occur in the initial cycles of vector generation and that as the number of vectors is increased the phenomenon appears to decay in a similar manner as shown in Fig. 8. This phenomenon is not as pronounced in the vector basis generated using fewer vectors per block, as indicated by the results shown in Fig. 8 for a block size of 1 vector. Asymptotic convergence of ϵ_u to zero can thereby be expected as the number of vectors is increased. A threshold value for ϵ_u to terminate the process of generation of vectors can not be assigned at this time, for its sensitivity to models with a more uniformly distributed mass is not known as well as the fact that the results presented are from only one loading condition.

A presentation of the norm ϵ_L associated with the translational components of the docking loading at time=1.93 seconds is given in Figs. 9 to 11. These results include the 60 and 90 Ritz vector basis (cases already discussed above, which contain the static residual identified as SR in the legends of Figs. 9 to 11) in addition to the 210 modal basis and a 60 Ritz vector basis without a static residual (identified as NSR in the legends). It can be observed that the Ritz vectors' basis can better represent the load than the modal vectors. This representation is improved as a static residual is added and further improved as the number of Ritz vectors in the basis is

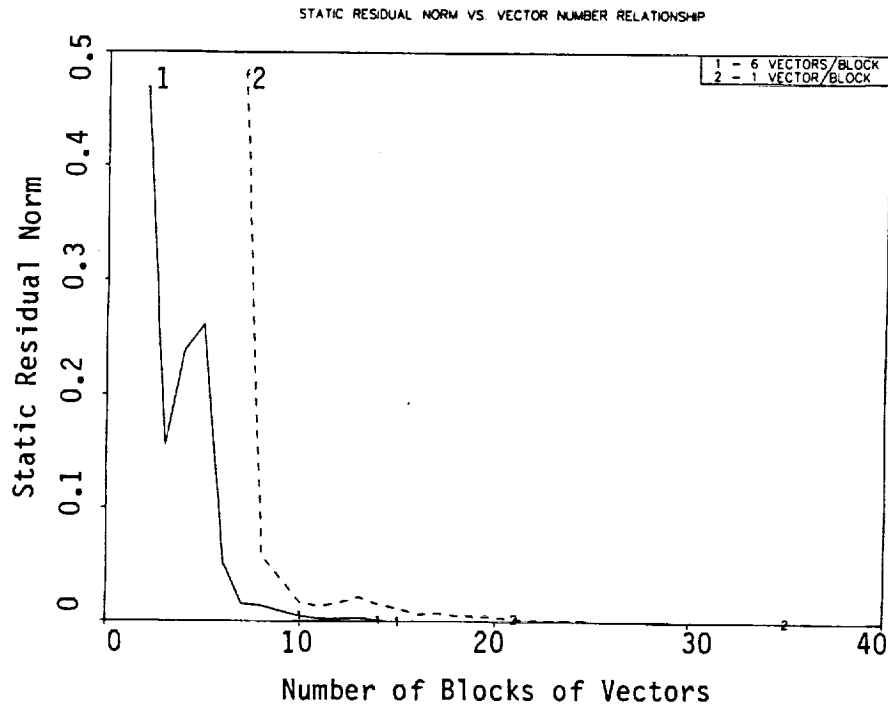


Fig. 8.- ϵ_u -Block Number Relationship, Block Sizes 1 and 6 Vectors.

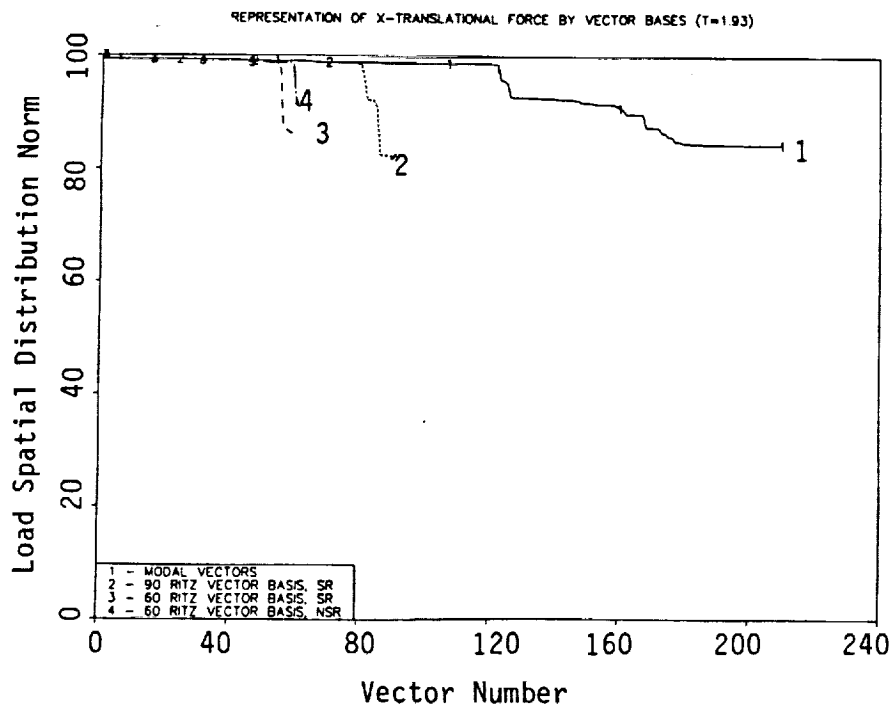


Fig. 9.- ϵ_L Norm for X-Translational Docking Force at Time 1.93 sec.

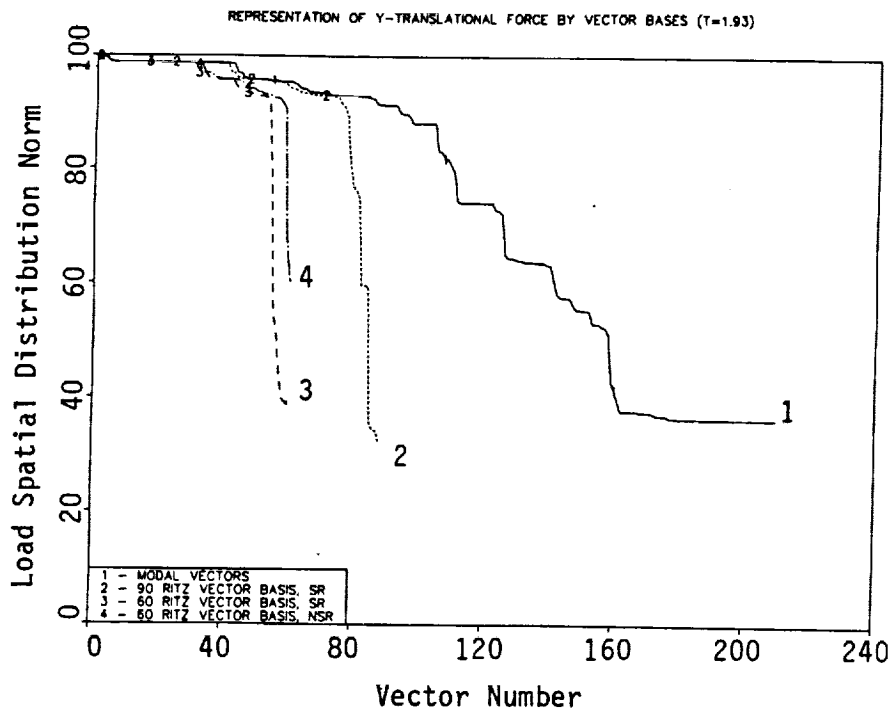


Fig. 10.- ϵ_L Norm for Y-Translational Docking Force at Time 1.93 sec.

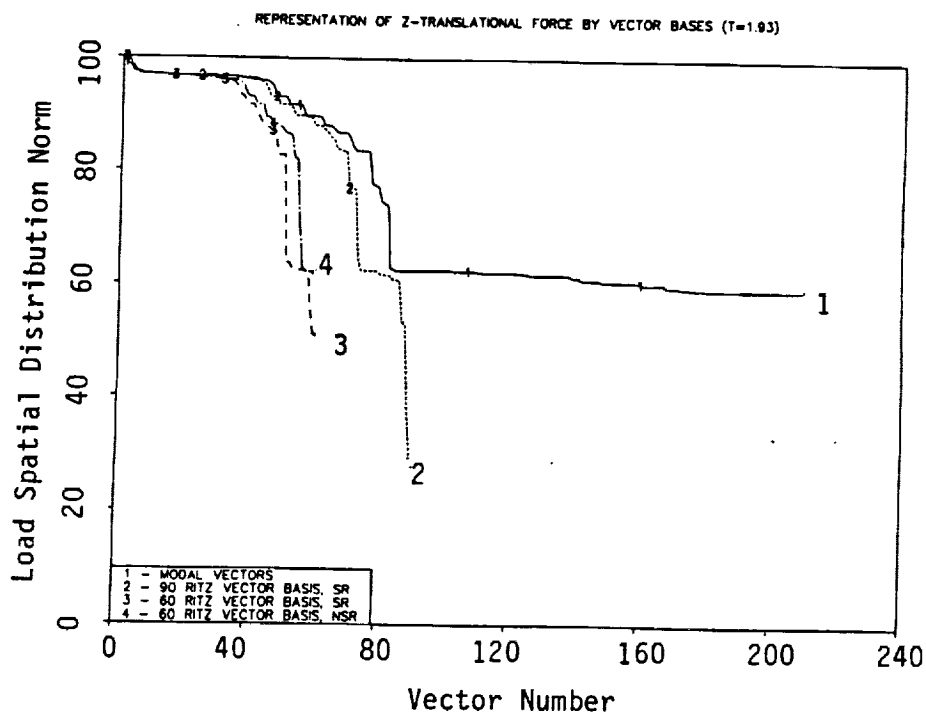


Fig. 11.- ϵ_L Norm for Z-Translational Docking Force at Time 1.93 sec.

increased. It can also be observed that representation of the loading by the 90 Ritz vector basis is as good (X- and Y-translation loadings) or better (Z-translational loading) compared to that of the 210 modal vector basis. The fact that fewer Ritz vectors were found to be required than modal vectors to achieve comparable results for ϵ_L gives an indication that the Ritz vector bases align more with the loading than the modal vector basis, such that the product of the Ritz vectors with the load vector is larger than the product using modal and the load vector. The participation of the Ritz modes oX in the dynamic response can therefore be expected to occur, as opposed to the possibility of having the vectors orthogonal to the loading as is sometimes the case in modal analysis. A close examination of Figs. 9 to 11 shows that most of the modal vectors' contribution to the load representation by its basis occurs in the mid frequency range (the vector number in Figs. 9 to 11 correspond to the ordering of the frequencies). Also, the modal vectors which are orthogonal to the loading are spread throughout the spectrum. This orthogonality is identified by the parts of the curve joining the values of ϵ_L where the slope is zero.

A comparison of the behavior of ϵ_u with that of ϵ_L indicates that although ϵ_u may become small, ϵ_L may not. Since convergence for accelerations was observed to require more Ritz vectors, more emphasis should be placed on ϵ_L for analysis in which the acceleration response is of prime interest when attempting to judge the quality of the vector basis for the selected loading conditions. This could perhaps be extended to situations where the Ritz vectors were generated based on a particular load case and there is the desire to use them for another load condition.

CONCLUSIONS

Based on the results of the preliminary analysis of the Space Station Freedom using the load-dependent Ritz vector algorithm presented herein, the following conclusions are noted:

1. Load-dependent Ritz vectors appear to provide accurate displacement solutions using fewer number of vectors than modal based vectors. The use of ϵ_u for criteria to stop the generation of the vector process appears to be promising.

2. More load-dependent Ritz vectors are required when computing accelerations. Criteria for judging the quality of the vector basis using ϵ_L needs to be more strict when acceleration response is of interest as opposed to displacements.
3. Load-dependent Ritz vectors have a broader frequency range than the equivalent number of modal vectors and are able to better represent the loading function. These features of the Ritz vectors are enhanced by including the static residual in the basis.

Areas which are suggested for future study include:

1. A study of the basic behavior and calibration of ϵ_u and ϵ_L for different load cases and structures, respectively.
2. Assessments of the reliability of the vector basis generated from a different load case than that used in the transient dynamic analysis and whether ϵ_L can provide a means of judging the vector basis quality in such cases.
3. The optimization of the algorithm to take advantage of the vectorization procedure by the CRAY X-MP/EA 464.

REFERENCES

1. Bathe, K., *Finite Element Procedures in Engineering Analysis*, Prentice-Hall, 1982.
2. Computerized Structure Analysis and Research Corporation, *CSA/NASTRAN Users Manual*, Vol. 1, 1988.
3. Nour-Omid, B. and R.W. Clough, *Block Lanczos Method for Dynamic Analysis of Structures*, Earthquake Engineering and Structural Dynamics, Vol. 13, pp.271-275, 1985.
4. Ricles, J., Leger, P., and L. Robayo, *Reducing Modal Truncation Error in the Wave Response Analysis of Offshore Structures*, Communications in Applied Numerical Methods, Vol. 6, pp 7-16, 1990.

59-52
N91-27P124
P 3

Mechanics, Impact Loads and EMG on the Space Shuttle Treadmill

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

Prepared by:
Academic Rank:

William G. Squires Ph.D.
Professor

University & Department: Dept of Biology Texas Lutheran College

NASA/JSC

Directorate: Space and Life Sciences

Division: Medical Sciences

Branch: SBRI

JSC Colleague: Mike Greensien, Ph. D.

Date Submitted: August 17, 1990

Contract Number: NGT-44-005-803

ABSTRACT

The ability of astronauts to egress the Shuttle, particularly during emergency conditions, is likely to be reduced following physiological adaptations in space.

It is well established that effective application of exercise counter measures requires the exercise be applied specifically. The problem is that objective scientific evidence is not available to validate the space shuttle treadmill with respect to its role in diminishing the deleterious effects of a prolonged exposure to the microgravity environment.

INTRODUCTION

The purpose of this study is to analyze locomotion and resulting impact loads during treadmill walking and running on orbit during STS missions. This data will be acquired with video motion tracking, force measurement instrumentation and electromyography (EMG). The data will be used to analyze the gait, to determine impact forces and to determine skeletal muscle responses, during extended 0 g conditions. These results will be used to determine appropriate exercise countermeasures prescriptions to counteract space adaptations in general and to prepare for the contingency of emergency egress in specific.

This study requires video recording astronaut performance during locomotion in zero g on the space shuttle treadmill mated to a six degrees of freedom force plate while wearing EMG telemetered electrodes

A total of twenty assigned astronauts (n=20) will be requested, five pilots and fifteen mission specialists, to participate in this study. Repeated measures on the same individual are necessary to avoid the large variance associated with comparisons between individuals.

This study will be conducted in three phases.

Phase 1. In the full fuselage trainer , crewmembers will be video recorded while performing locomotion tasks on the instrumented treadmill while wearing EMG telemetered electrodes.

Phase 2. During parabolic flight on the KC 135, crewmembers will be video recorded performing locomotion tasks on the instrumented treadmill while wearing EMG telemetered electrodes.

Phase 3. On orbit, crewmembers will be video recorded while performing locomotion tasks on the instrumented treadmill while wearing EMG telemetered electrodes.

An accurate measurement of the gait patterning, locomotion impact loads and EMG data can provide real data into the system for further exercise countermeasure prescriptions.

510-90
20225
N91-27113
P-13

SOLUTIONS TO PROBLEMS OF WEATHERING IN ANTARCTIC EUCRITES

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

A0659909

Prepared By:	Melissa M. Strait, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	Alma College Department of Chemistry Alma, Michigan 48801

NASA/JSC

Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch:	Planetary Science
JSC Colleague:	Marilyn M. Lindstrom, Ph.D.
Date Submitted:	August 10, 1990
Contract Number:	NGT-44-005-803

ABSTRACT

The number of samples available for studies of asteroidal differentiation and early Solar System history has been dramatically increased with the discovery of meteorites in Antarctica. However, it has been recently shown that significant differences in minor and trace element abundances between Antarctic and non-Antarctic eucrites are probably due to weathering (Mittlefehldt and Lindstrom, 1990). This alteration is thought to occur in interstitial material leaving primary silicate phases intact and suitable for petrologic modelling.

This study performed neutron activation analysis for major and trace elements on a suite of eucrites from both Antarctic and non-antarctic sources. The chemistry was examined to see if there was an easy way to distinguish Antarctic eucrites that had been disturbed in their trace element systematics from those that had normal abundances relative to non-Antarctic eucrites. There was no simple correlation found, and identifying the disturbed meteorites still remains a problem.

In addition, a set of mineral separates from an eucrite studied that had been studied earlier were analyzed. The separates were acid-washed to remove any interstitial material or contamination before they were analyzed. The results showed no abnormalities in the chemistry and provides a possible way to use Antarctic eucrites that have been disturbed in modelling of the eucrite parent body.

INTRODUCTION

The eucrite association (eucrites, howardites, diogenites and mesosiderites) represents an occurrence of planetary volcanism from the smaller bodies of the Solar System. This meteorite association is studied to understand chemical and genetic relationships among the meteorite groups and to look at processes that occurred 4.5 billion years ago in the Solar System.

The eucrites are pyroxene-plagioclase achondrites, the diogenites are hypersthene achondrites, the howardites are polymict breccias of eucritic and diogenitic material and the mesosiderites are stony irons with silicates of a composition similar to the eucrites. A spectrum of chemical and mineralogical compositions tie together the eucrites, howardites and diogenites (Figure 1). At one time all four classes of meteorites were considered to come from one parent body, possibly even the Moon (Duke and Silver, 1967). The return of lunar samples eliminated the Moon as a parent body, and recent work on chemistry and mineralogy of the mesosiderites has shown that while they may form in a similar process to the howardite-eucrite-diogenite grouping, they come from a separate parent body (Mittlefehldt, 1990).

This study will look specifically at the eucrites to examine the chemical effects of weathering and to further ongoing studies of asteroidal differentiation and early Solar System history. Eucrites have textures that resemble terrestrial basalts and are modelled in an analogous manner. They consist of calcium-poor pyroxenes and calcium-rich plagioclases, with the pyroxene generally dominant over plagioclase. In addition, the eucrites contain accessory amounts of olivine, tridymite, quartz, cristobalite, chromite, magnetite, ilmenite and phosphates, such as apatite and whitlockite, as well as minor amounts of metal and troilite (Duke and Silver, 1967).

The eucrites can be divided into several classes based on structural and chemical composition. Structurally, the meteorites are divided into monomict and polymict brecciated meteorites. Classically, eucrites were monomict and howardites were polymict. The new influx of meteorites from Antarctica has necessitated the addition of a mineralogical-chemical criteria which adds a group called polymict eucrites. These have a polymict structure, but eucritic compositions (Delaney, et al., 1983). The polymict eucrites contain clasts that represent a wide range of igneous lithologies, including eucrites, two-pyroxene mafic rocks, pigeonite cumulate eucrites, alkali-enriched mafic rocks, diogenites, fayalite rich clasts and feldspar cumulate eucrites. (Delaney, et al., 1984).

There are two models for the formation of the eucrites. Both models depend upon different interpretations of the same set of samples and analyses. Mason (1962) proposed that the eucrites formed by the complete melting of an asteroid and that the meteorites might be related as successive members of a fractional crystallization sequence from a melt of approximately chondritic composition. The eucrites were derived from a thin surface crust, while the core consisted of pallasitic material (another type of stony-iron meteorite) surrounded by a diogenitic mantle (Mason, 1967). Stolper (1977) proposed partial melting of the eucrite parent body in successive stages that produced the different types of basaltic achondrites. The different groupings of meteorites were produced when melting occurred in the source region from which earlier melts had been removed.

In 1969 a important event happened in the world of meteoritics: meteorites were discovered in abundance on the ice in Antarctica. Prior to this about 2000 meteorites had been identified in the 250 years since meteorites were recognized to have come from outside the Earth. Since 1969, approximately 12000 meteorite samples have been recovered from the ice in Antarctica by teams of scientists sent from the US and Japan. Samples from Antarctica are individual rocks that have been picked up from the ice. Some of these were members of the same meteorite that shattered before or upon impact with the ice. Chemical and petrographic analysis allows these to be paired. This means that the number of samples from Antarctica is much larger actual number of meteorites represented.

Meteorites from Antarctica have extended the amount of material available on which to do meteorite research, as well as filled in gaps in existing suites of meteorites. In addition, new types of

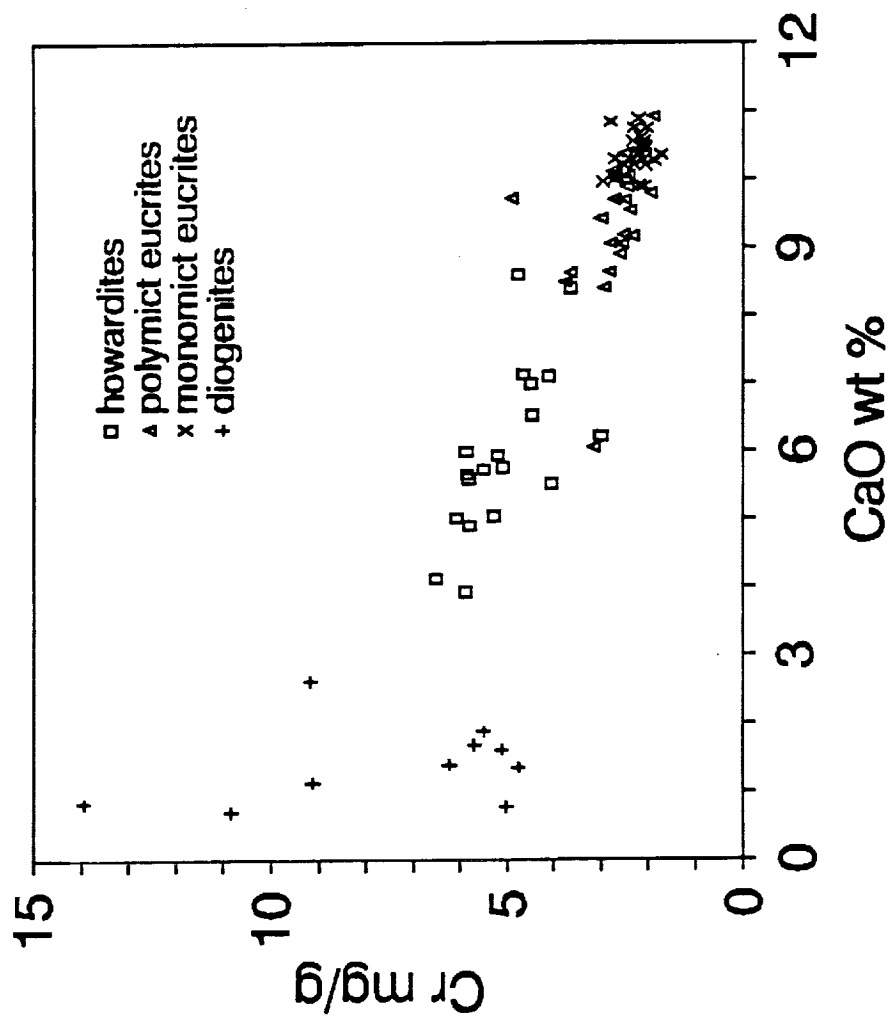


Figure 1. The chemical relationship between the types of basaltic achondrites is demonstrated. Note the close overlap of the monomict and polymict eucrites.

extraterrestrial material have been discovered, including meteorites that originate from a known source - the Moon.

About 4300 meteorite samples have been returned from Antarctica by US scientists, of which ~80 have been classified as basaltic achondrites (Score and Lindstrom, 1990). This includes 56 samples of eucrites which represent approximately 28 meteorites, as well as 17 howardite samples and 8 diogenite samples.

Meteorites are traditionally divided into two types - falls and finds. "Falls" are meteorites that have been recovered from observed meteorite falls. Therefore their residence time upon the earth is well documented. Finds are stones that have been found and subsequently recognized as meteoritic. Their residence time upon the earth is unknown, although there are ways that this can be approximated (Schultz, 1990). Most prior meteorite work has been done on material from falls. This reduces the chance that observations are the result of interactions that occurred upon the earth rather than the processes that are of interest - those that occurred during the formation and evolution of the solar system.

With the wealth of material coming from Antarctica, this strategy has changed. Meteorites that are recovered from the ice have spent from an average of about 230,000 years up to as long as 1,000,000 years immersed in the Antarctic icesheet (Schultz, 1990). Despite the low temperatures, it has come to be recognized that chemical reactions are occurring that alter the meteorites from their pristine state (Gooding, 1990). These meteorites must be carefully examined to look for the effects of weathering prior to using them in any studies of their origin.

Prior work with the Antarctic eucrites has led to the definition of two groupings based on the abundances of the REE: those with normal REE patterns, and those with abnormal patterns, mostly apparent enrichments in Ce and Eu (Mittlefehldt and Lindstrom, 1990). This is an important observation. The REE are used extensively to model igneous materials and abnormal abundances due to causes other than igneous processes must be examined.

The REE are incompatible in most minerals that crystallize in a melt. They tend to act as a group and not be fractionated from one another easily. There are slight differences in their size, however, and this leads to variable incorporation into mineral phases and consequent known fractionations. The fractionations are smooth functions of atomic number. An analysis of the REE contents of a suite of related geologic materials provides evidence for the geochemical history of these samples. The only common exception to the smooth fractionation is Eu because under certain oxidizing-reducing (redox) conditions, the element will exist in the +2 oxidation state rather than the more normal +3 state exhibited by the other REE when the conditions are reducing enough. In the +2 state, it has a different ionic radius, and is incorporated into different mineral phases than the other REE.

In all cases, the REE are considered incompatible, and do not readily enter into the crystallizing minerals until the late stages when they are higher in abundance in the melt. As the melt continues to crystallize, the solids will be progressively more concentrated in REE as the REE are concentrated in the melt. Therefore, early and late materials from a single melt body can be determined by their REE abundances. The size of the Eu anomaly gives clues as to how much plagioclase has been previously crystallized from the melt.

Abnormalities in the REE pattern, one or more elements breaking the smooth pattern, are not readily explained in igneous materials. In chondritic materials, which are presumed to condense from the solar nebula, abnormalities are sometimes observed which can be attributed to localized redox conditions in the solar nebula. In an achondritic material, however, these abnormalities will have been reset by melting and recrystallization.

It has recently been proposed (Mittlefehldt and Lindstrom, 1990) that the abnormalities observed in some Antarctic eucrites are caused by terrestrial weathering that occurs in the Antarctic ice. The primary REE minerals in the eucrites are the calcium phosphates. The phosphates are readily susceptible to dissolution in slightly acidic solutions. It was proposed that the phosphates dissolved

during the residence of the meteorite in Antarctica. When the phosphates dissolve and are removed from the meteorite, they carry with them a large part of the REE budget of the meteorite. Because Eu has been preferentially crystallized in the plagioclase, which is much less susceptible to alteration, Eu is retained in the meteorite. It is proposed that the Ce anomaly arises because it is oxidized to the +4 state by the acidic conditions that leach the apatite. Ce + 4 forms an insoluble Ce oxide which precipitates and remains behind in the meteorite. Therefore, the abnormalities are not enrichments of Ce and Eu, but depletions of the other REE.

This doesn't appear to happen in all cases. There are Antarctic eucrites with normal REE patterns that fall within the range of patterns exhibited by non-Antarctic eucrites. However, the presence of these abnormal Antarctic eucrites presents problems for workers who model the origins and evolution of the eucrites using REE data. Using abnormal eucrites will lead to erroneous results in models. This study had a two-fold purpose: first to enable abnormal eucrites to be readily recognized and second to look at ways to use abnormal eucrites in petrogenetic models.

SAMPLES

In the first case a suite of eucrites from a variety of locales was analyzed. Included in the set were five non-Antarctic meteorites to enable comparisons to be made with stones that have not undergone Antarctic exposure. All five non-Antarctic meteorites were falls. Antarctic meteorites were sampled in a variety of ways. Several groupings of clast and matrix samples were taken, as well as a paired interior/exterior samples. This latter was done with the assumption that weathering is more extensive on the exposed outer surface of a rock than in the enclosed interior of a rock. The latter assumption is not always true, as cracks and fissures in the stone allow water to enter the supposedly pristine interior of the rock. Care must be taken to examine all samples for evidence of weathering, regardless of whether it is an interior or an exterior sample. The easiest indication of weathering is ferrous rust, which can result from the oxidation of metal and troilite in the sample. Unfortunately, in the case of eucrites, these minerals are rare and the ubiquitous weathering is the alteration of silicates to produce a silic rust (Gooding, 1990), which does not have the characteristic "rusty" color of weathered iron.

The second part of the study attempted to find a way to use the abnormal Antarctic eucrites. An abnormal sample, LEW85300,57, examined in an earlier study (Mittlefehdt and Lindstrom, 1990) was utilized. The sample was separated into its constituent minerals. If the weathering truly affected primarily the phosphates, the primary pyroxene and plagioclase should be minimally affected. Minerals separates were performed in two different ways: magnetic separation and heavy liquid density separations.

The separates were washed with a series of different acids to remove phosphates, cerium oxide, and other interstitial material from the pyroxene and plagioclase. This should remove the phases that caused the abnormal REE patterns in the whole rocks.

EXPERIMENTAL

All samples were described, obvious weathered material was removed (this was primarily rusty appearing material), the samples crushed in a diatomite mortar and pestle and packaged in silica tubes for irradiation. In addition to the meteorite samples, four standard samples were included to determine concentrations, as well as three standard materials included as system monitors. Irradiation was performed at the University of Missouri Research Reactor at a thermal neutron flux of $\sim 5 \times 10^{13}$ n-cm⁻²-sec⁻¹. The samples were counted three times in the JSC neutron activation facility to obtain data on a number of nuclides with varying half-lives. The spectra were processed in an updated version of the TEABAGS program (Lindstrom and Korotev, 1982) and further processed using undocumented

software to perform various corrections and determine concentrations. The final results are available from the author at the listed address.

RESULTS

The results were examined in a variety of ways. Antarctic samples were compared to the non-Antarctic meteorites analyzed, clasts were compared to matrix and interior samples were compared to exterior samples. Observations of the sample sets are summarized below and followed by a discussion of the entire suite of samples.

LEW88005 - Three splits of this sample were analyzed, two clasts (,13 and ,16) and one matrix (,22) sample. Split ,16 showed little weathering, while split ,13 had obvious weathering, mostly non-ferrous. The matrix had slight amounts of weathering. The two clasts were different in their chemistries, especially when compared to the matrix. ,16 tended to track the matrix more closely than did the more obviously weathered ,13. All of the splits exhibited Ce anomalies; the two clast samples exhibited Eu anomalies, but the matrix sample did not. Barium was enriched in the clast relative to the matrix, and Cr and Co depleted; in all cases variations in split ,16 were more pronounced than split ,13. Split ,13 was depleted in the REE relative to the matrix, while split ,16 was not.

EET87542 - Three splits were also analyzed from EET87542. The two matrix (,11 and ,16) samples and one clast (,16) sample were all remarkably consistent in their chemistries. The only exceptions were U, which was enhanced and K and Na which were slightly depleted in the clast relative to the matrix. None of the samples exhibited Ce anomalies. All of the samples were described as having small rusty patches pervasive over the surface, larger patches of which were removed.

HOW88401 - Five samples were taken from HOW88401, both clast - matrix pairs and an interior - exterior pair. Splits ,13 and ,15 were compared as an interior/exterior pair, while clasts splits ,9 ,13 and ,21 were compared to matrix split ,25. This meteorite showed more variety in its chemistry than the other clast - matrix pairings. There was a wide spread of REE abundances, as well as a variety of patterns. Splits ,13 and ,25 showed the Ce anomaly indicative of abnormal eucrites. With the exception of split ,15, the exterior sample, none of the samples were originally described as having any obvious weathering. Split ,21 is depleted in all of the measured elements relative to the matrix. Split ,9 was enriched in all elements except Fe, Sc, Cr, Co and Eu, which are the same as the matrix. Split ,13 was depleted in all elements relative to the matrix except Ca, Fe, Sc, Cr and Co. Split ,13 had a large piece of chromite, which is shown by its enriched Cr. The matrix pattern was flat at 7 times the cosmic abundance of the REE with no Eu anomaly, but a positive Ce anomaly.

The interior/exterior comparison leads to some interesting observations. First, the interior is depleted in all elements relative to the exterior except for Fe, Cr and Sc. Fe and Cr are enriched in the interior, while Sc is not changed. The REE in the exterior sample are the most enriched of all the splits from this sample, has no Ce anomaly, but does have a pronounced negative Eu anomaly.

Other Eucrites - Of the remaining five samples, three exhibited Ce anomalies and two did not. One of the latter samples (EET85548,12) is apparently a plagioclase cumulate with low REE abundances. The Ce abundance was below the detection limit, so it was not possible to tell if there was a Ce anomaly. This sample was observed to have slight weathering. Of the other four samples, two were described as having little to no weathering (LEW85353,6 and EET87520,14), while the other two exhibited mild weathering (LEW87026,11 and RKPA80204,16). The samples were all whole rock, and the abundances of elements other than the REE tended to fall within ranges established by non-Antarctic eucrites with the exception of EET85548,12 the plagioclase cumulate.

LEW85300,57 - In this study only mineral separates were analyzed of LEW85300,57. In a prior study (Mittlefehldt and Lindstrom, 1990), a suite of whole rock samples from the LEW8530x suite were analyzed. The ,57 split exhibited extensively disturbed REE systematics. It was chosen for this study because of this. Four mineral separates were analyzed, of which data was obtained for three. The two

pyroxene samples were almost identical, except for Eu, which was more depleted in the heavy density pyroxene separate ($3.55 < d < 3.7$) than in the pyroxene obtained by magnetic separation. The pattern obtained was a typical pyroxene pattern with depleted LREE relative to HREE and a negative Eu anomaly. The plagioclase pattern obtained from the magnetic separation was also a typical pattern for plagioclase: LREE enriched and a strong positive Eu anomaly. None of these samples exhibited a Ce anomaly.

DISCUSSION

In general, the Antarctic meteorites fall in the same range of abundances that the non-Antarctic eucrites do. The exceptions were generally readily explained, such as being extensively weathered (HOW88401,15) or a different lithology (the cumulate eucrite EET87548,12).

The suite of samples as a whole was examined to determine if there was a simple way to distinguish abnormal from normal Antarctic eucrites. Abnormal eucrites have been defined as samples which have more than a 10% fractionation of La from Ce. Seven of the sixteen samples exhibited Ce anomalies, however only two of them show greater than 20% fractionation. This data set, therefore is not a particularly good one for looking at differentiating normal from abnormal eucrites (Figure 2).

Easy, obvious ways of differentiating the two groups are not particularly useful. Weathering, identified by naked eye or using a binocular microscope, whether ferrous or silic, shows no apparent correlation between the amount of weathering observed on the meteorite and the presence of a Ce anomaly. Of the seven samples which had a Ce anomaly, four of them had no obvious weathering. The most heavily weathered sample, the exterior one, had no Ce anomaly.

Even when this data set is combined with earlier data collected to look at the Ce anomaly problem, no simple result comes out that unequivocally sets apart the abnormal eucrites. Previous work has shown that chondrite normalized Ce/La and Eu/Sm deviate from a chondritic ratio of one more extensively in the Antarctic meteorites. This is not clearly demonstrated in this limited data set, but the one sample in which Eu/Sm deviates from chondritic is one of the abnormal samples. The other samples fall within a typical range as established by non-Antarctic eucrites. It was attempted to use Hf, another incompatible element in the same fashion, but similar results were obtained (Figure 3).

The interior - exterior pair in this study showed depletions for all elements except iron and chromium in the interior of the sample relative to the exterior (Figure 4). Apparently the materials being removed contain Ca, Na, and most of the trace elements. This would correspond to phosphates and other interstitial material. Pyroxene, which contains most of the iron, and chromite, which contains the chromium, are less susceptible to weathering and are not removed.

Leached materials appear to be removed from the interior and deposited on the outside of HOW88401. However, this is not always the case. Other samples from the prior study show variable results for interior/exterior couples. The materials leached from the interior are sometimes deposited on the exterior of the sample, but in other cases, are completely removed from the sample. It would be interesting to analyze the ice in contact with an abnormal eucrite and see if the leached elements end up in the ice.

Finally, the attempt to see if abnormal eucrites potentially could be used for geochemical modelling seems to have been successful. The mineral separates from LEW85300,57, a decidedly disturbed eucrite, show normal REE patterns with no evidence of the disturbance present in the whole rock. Comparisons between the results obtained in this study and earlier studies on mineral separates from non-Antarctic eucrites show good agreement (Duke and Silver, 1967; Schnetzler and Philpotts, 1969).

According to our results, the Ce anomaly does arise in the interstitial material as hypothesized (Mittlefehldt and Lindstrom, 1990) and not in the major minerals. This result does not agree with an earlier study by Heavilon and Crozaz, 1990 using in situ secondary ion mass spectrometry. They looked

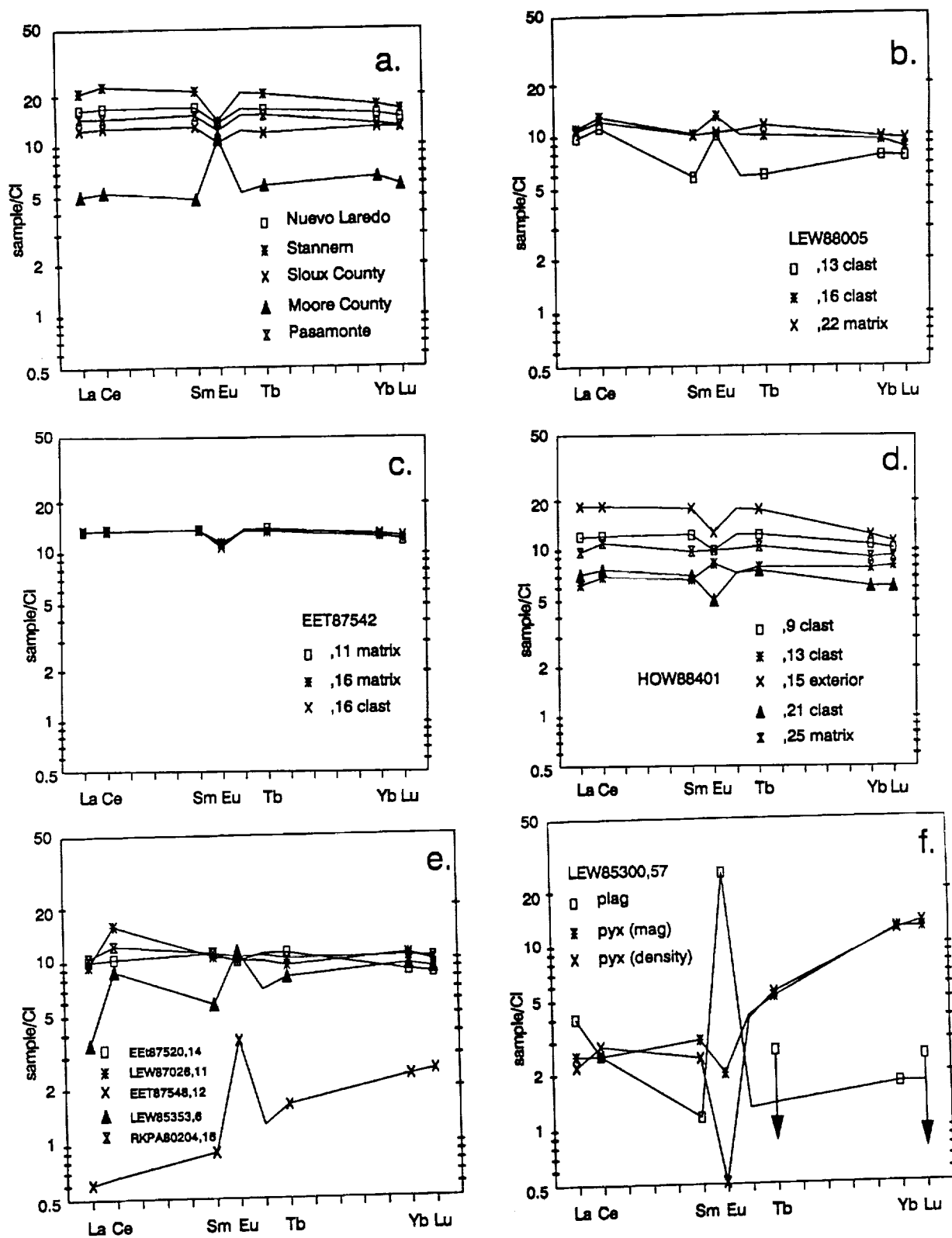


Figure 2.- The REE patterns for the eucrites analyzed in this study.
a) The non-Antarctic eucrites, b) LEW88005, c) EET87542, d) HOW88401,
e) a set of whole rock samples, f) mineral separates.

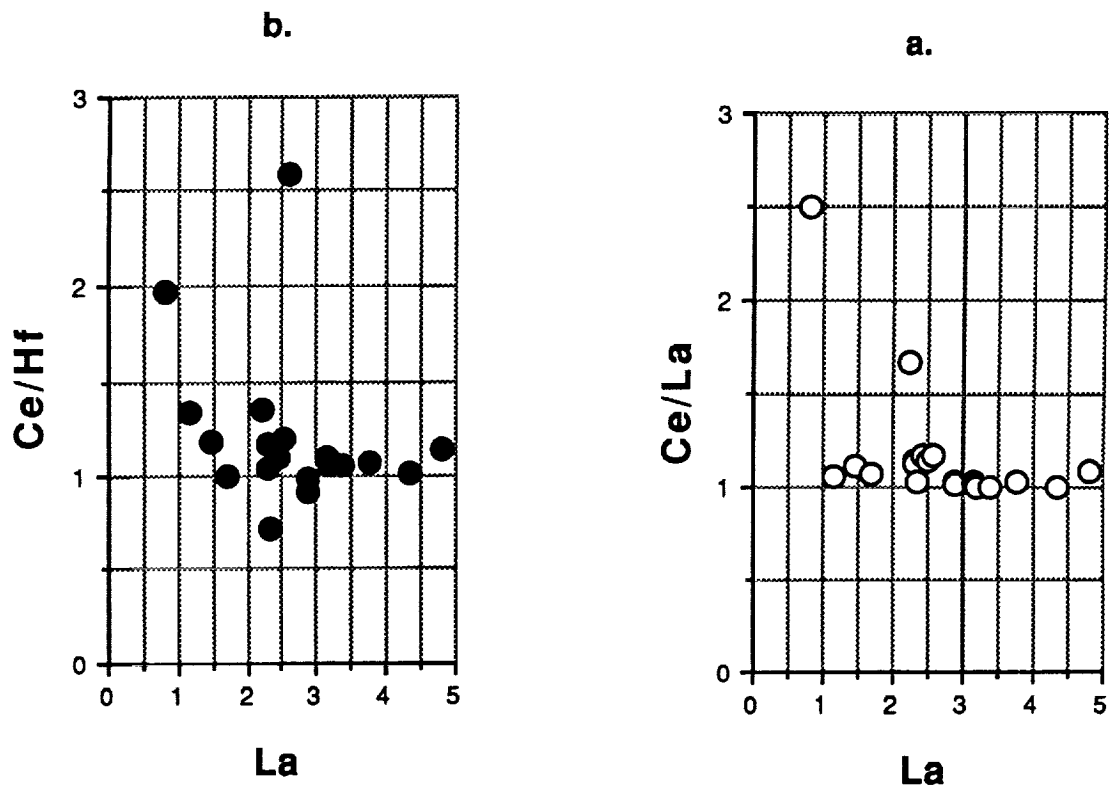


Figure 3.- Ratios of trace elements used to demonstrate differences between Antarctic and non-Antarctic meteorites. a. Ce/La b. Ce/Hf.

Interior/Exterior Comparison

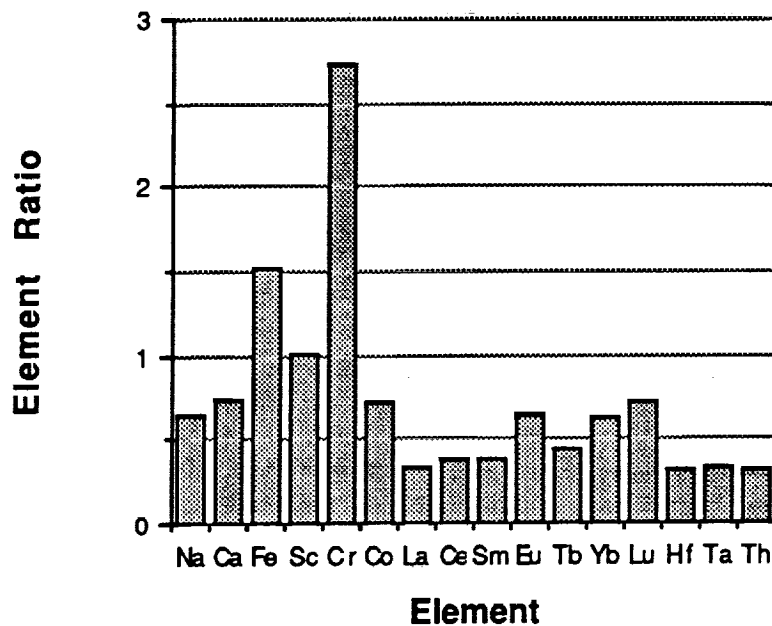


Figure 4.- Demonstration of the migration of elements between the interior and exterior of HOW88401.

at thin sections of the same meteorite, LEW85300, and found Ce anomalies present in pyroxene. Both positive and negative Ce anomalies were present, including adjacent spots with opposite anomalies or even no anomaly. The anomaly was randomly distributed. If this is real, the small scale variations in individual grains may be averaged out by the larger amounts of material analyzed using instrumental neutron activation analysis, as in this study. Or, the variable results may be due to interstitial material smeared on the surface of the thin section during preparation. In our study we removed these types of potential contaminants with the acid leaches. Further work needs to be done to resolve this problem.

Potentially, however, it appears that if a sample could be readily identified as abnormal, it could still be used for doing geochemical modelling if care is taken in handling the materials, i.e., using minerals separates.

CONCLUSION

In conclusion, this study did not establish an easy way to distinguish abnormal from normal Antarctic eucrites. However, if the sample is identified as abnormal, it can be used for modelling if mineral separates of the pyroxenes and plagioclases are used rather than whole rock data. This will require extra work on the part of the investigator, but will allow meteorites from Antarctica to be used in investigations of the evolution of the Solar System.

ACKNOWLEDGEMENT

I would like to acknowledge and thank David W. Mittlefehldt of Lockheed ESC for his help and support in this project. He re-introduced me to the neutron activation lab, and he provided access to his vast knowledge of the eucrites and other basaltic achondrites.

REFERENCES

- Delaney, J.S., Prinz, M. and Takeda, H.: The Polymict Eucrites. *Jour. Geophys. Res.*, vol. 89, 1984, pp. c251-c288.
- Delaney, J.S., Takeda, H., Prinz, M., Nehru, C.E. and Harlow, G.E.: The Nomenclature of Polymict Basaltic Achondrites. *Meteoritics*, vol. 18, 1983, pp. 103-111.
- Duke, M.B. and Silver, L.T.: Petrology of eucrites, howardites and mesosiderites. *Geochim. Cosmochim. Acta*, vol. 31, 1967, pp. 1637-1665.
- Gooding, J.L.: Clay-mineraloid weathering in Antarctic meteorites. *Geochim. Cosmochim. Acta*, vol. 50, 1986, pp. 2215-2223.
- Heavilon, C.F. and Crozaz G.: Ce anomalies in the Antarctic eucrite LEW85300. *Lunar Planet. Sci.*, vol. XXI, 1990, pp. 487-488.
- Lindstrom, D.J. and Korotev, R.L.: TEABAGS: computer programs for instrumental neutron activation analysis. *J. Radioanal. Chem.*, vol. 70, 1982, pp. 439-458.
- Mason, B.: *Meteorites*. John Wiley and Sons, 1962.
- Mason, B.: Meteorites. *Amer. Scient.*, vol 51, 1967, pp.429-455.
- Mittlefehldt, D.W.: Petrographic and chemical characterization of igneous lithic clasts from mesosiderites and howardites and comparison with eucrites and diogenites. *Geochim. Cosmochim. Acta*, vol. 43, 1979, pp. 1917-1935.
- Mittlefehldt, D.W. and Lindstrom, M.M.: Generation of abnormal trace element abundances in Antarctic eucrites by weathering processes. *Geochim. Cosmochim. Acta*, 1990, in press.
- Schnetzler, C.C. and Philpotts, J.A.: Genesis of the calcium-rich achondrites in light of rare-earth and barium concentration. In *Meteorite Research*, ed. P. M. Millman, Reidel, pp. 206-216.
- Schultz, L.: Terrestrial ages and weathering of Antarctic meteorites. LPI Tech. Rpt. 90-03, 1990, pp. 56-59.
- Score, R. and Lindstrom, M.M.: Guide to the U.S. Collection of Antarctic Meteorites 1976-1988. *Ant. Met. Newslet.*, vol. 13, 1990.
- Stolper, E.: Experimental petrology of eucritic meteorites. *Geochim. Cosmochim. Acta*, vol 41, 1977, pp.587-611.

511-52
20226
P-14
N91-27114

ASSESSMENT OF THE PHARMACODYNAMICS OF INTRANASAL, INTRAVENOUS
AND ORAL SCOPOLAMINE

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1990

Johnson Space Center

PM 162291

Prepared By:	Karen J. Tietze, Pharm.D.
Academic Rank:	Associate Professor
University & Department:	Philadelphia College of Pharmacy and Science Department of Pharmacy Practice/ Pharmacy Administration Philadelphia, PA 19104-4495
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Medical Sciences
Branch:	Biomedical Operations and Research
JSC Colleague:	Lakshmi Putcha, Ph.D.
Date Submitted:	August 10, 1990
Contract Number:	NGT-44-005-803

ABSTRACT

Space motion sickness is an important issue in the space medical sciences program. Although scopolamine has been documented to be an effective antimotion sickness agent, limited information is available regarding the pharmacodynamics of scopolamine and the relationship between the pharmacodynamics and the pharmacokinetics of the drug. One of the objectives of the ongoing clinical experimental protocol "Pharmacokinetics of Intranasal Scopolamine in Normal Subjects" is to evaluate the pharmacodynamics of scopolamine using salivary flow rate and pH profiles and cognitive performance tests as pharmacodynamic parameters.

Normal volunteers collected saliva and performed the NTI Multiresource Performance Battery test at designated time intervals to establish control salivary flow rates, salivary pH profiles, and the characteristics of the learning curve for the performance program under normal conditions. Salivary flow ranged between 50 to 60 ml/hr with pH ranging between 6.5 to 7 over a 2 hour period. Preliminary assessment of the performance data from the first few volunteers suggests that the duration of the learning curve is short reaching a plateau after two test sessions for task one and five to eight test sessions for task two.

In the clinical part of the study, saliva samples and performance test scores are collected from healthy nonsmoking subjects after receiving a single 0.4 mg dose of either intranasal, intravenous, or oral scopolamine. Both salivary flow rate and pH decreased as a function of time after administration of all the three dosage forms. The pharmacodynamic effect as indicated by salivary flow rate and pH was the greatest after intravenous administration and the least after oral administration. Preliminary analysis of the performance parameters for a limited number of subjects indicates that there is no significant change in task one or task two scores following the administration of any of the three dosage forms. However, the mean correct response times for tasks following the transition from one task to another appear to be slightly longer during the first few hours following scopolamine administration.

Preliminary evaluation of this limited data suggests that salivary flow rate and pH are good pharmacodynamic indicators for scopolamine, and that intranasal scopolamine produces a reliable and consistent pharmacodynamic effect. Limited performance data collected so far suggest that scopolamine does not decrease cognitive function as measured with the NTI Multiresource Performance Battery. Results from the study may be useful in identifying and selecting effective dosage forms for the management of SMS.

INTRODUCTION

Space motion sickness (SMS) is a major concern in the space medical sciences program. Approximately 50-60% of astronauts experience SMS, which usually affects astronauts during the first three days of flight and may be severe enough to compromise the operational performance of the astronaut (1). Use of antimotion sickness drugs in flight has been limited by drug ineffectiveness or by undesirable side effects that impair the operational performance of the astronaut. Scopolamine, an anticholinergic drug, has been documented to be an effective antimotion sickness agent (2,3). It has been suggested that scopolamine interrupts the acetylcholine-dependent neurotransmission of information from the vestibular organs to the vomiting centers of the brainstem (4,5). At present, scopolamine (0.4 mg) in combination with dextroamphetamine (5 mg) is administered orally as a capsule for the prophylaxis and treatment of SMS. An important side effect of operational significance induced by scopolamine is the impairment of cognitive performance. Another side effect that is not of clinical importance but can be used as a pharmacodynamic indicator is inhibition of salivary secretion.

Although a veteran drug, the pharmacodynamics of scopolamine and the relationship between the pharmacokinetics and the pharmacodynamics of the drug have not been well characterized. Several investigators have documented that oral scopolamine impairs cognitive performance; the techniques used are complex and time-consuming (6-12). Identification of a sensitive and convenient method for the assessment of drug-induced performance changes would be helpful in establishing the selection criteria for drugs used for the prophylaxis and treatment of SMS and also for the pharmacodynamic assessment of these drugs during space flight. The project therefore was designed to 1) establish salivary parameters in control subjects using the Sarstedt Salivette system, 2) to determine cognitive performance in control subjects using the NTI Multiresource Performance Battery, and 3) to compare the effect of a 0.4 mg dose of intranasal, intravenous, and oral scopolamine on these parameters.

MATERIALS AND METHODS

Scopolamine Study

A conventional clinical IND protocol was designed to evaluate the pharmacokinetics and bioavailability of the three dosage forms. Healthy nonsmoking male subjects ranging between 18 to 40 years of age are participating in the study. All subjects passed an Air Force Class III physical examination, including hematology, electrocardiogram, urinalysis, and biochemistry tests before admission to the study. The study protocol was approved by the human research review committees of NASA-JSC and the St. John Hospital. Informed consent was obtained from all the subjects prior to the study.

The study consists of an intranasal, an intravenous, and an oral treatment phase. A 0.4 mg dose is used in all three treatment phases. Each 24 hour treatment phase is separated by at least a two week period. The treatments are administered in a predetermined randomized crossover design. Oral scopolamine was customed manufactured by AC Engle & Co, Houston, Texas, intravenous scopolamine dosage form was procured from a pharmacy, and intranasal scopolamine was customed-manufactured by the Pharmaceutics Department of the University of Houston, Houston, Texas. Subjects are admitted to the hospital the night before each study period. Each subject is given written and oral instructions for the Multiresource Performance Battery computer program and trained until they are comfortable with the computer keyboard and the program. Mixed saliva is collected over a two minute interval using the Sarstedt Salivette system (a cotton swab collection system) before and at 5, 10, 15, 20, 30, 45, 60, 90 minutes and 2, 3, 4, 5, 6, 8, 10, and 12 hours after dosing. The Multiresource Performance Battery test is taken pre-dose and at 1, 2, 3, 4, 5, 6, 8, 10, and 12 hours post-dose.

The investigational Multiresource Performance Battery computer program was obtained from NTI, Inc. The program runs on a portable personal computer and is designed to test the higher cognitive functions most likely to be affected by centrally active drugs. The program consists of two timed tasks presented pseudosimultaneously with an attention allocation indicator. Both tasks appear on the screen simultaneously with two bars of differing height at the bottom center of the screen. The active task is identified as the task on the side of the screen with the highest bar. The math task appears on the left of the screen and consists of a two step (addition, subtraction, or combination) single digit math problem. The key "z" is pressed if the answer is less than five and the key "v" is pressed if the answer is greater than five. The memory task appears on the right of the screen and consists of two single digit numbers separated by a line. The task is to determine if the numerator of the current active task matches the denominator of the prior active task. The key "m" is pressed if the numbers match and the key "/" is pressed if the numbers do not match. The test battery consists of 100 randomly generated tasks with an approximately equal distribution between the two tasks. Each 100 task test battery is preceded by a randomly generated 20 task warm-up battery that may be repeated until the subject no longer feels his performance is improving. The program allows for a maximum of five seconds for a response before randomly changing to a new test in the same task or to the alternate task. Performance parameters include the percent correct, incorrect and timed out and the mean correct response times for both tasks as well as the percent correct, incorrect, and timed out and the mean correct response times for the transitions to each task and for the total transitions.

Control Data Collection

Because the NTI Multiresource Performance Battery is a new investigational program, no data exists on the learning curve with the program. Preliminary analysis of data from the first few subjects suggested a learning curve interval which may mask the drug effect. Another contributing factor of variability may be the time of day (i.e. morning vs. afternoon). To address these issues, ten normal volunteers were selected for twice daily testing for ten days. Each subject received verbal and written instructions for the Multiresource Performance Battery and was then tested twice daily (in the morning and afternoon) for a total of 20 tests over a three week period. The performance parameters (percent correct, incorrect and timed out; mean correct response times; and transition percents and response times) for each task will be analyzed and the learning curve will be characterized.

To evaluate whether the repeated collection of saliva with the Salivette system during the first hour of the study would alter salivary flow dynamics, four volunteers repeated the saliva sampling protocol for the clinical study. Salivary volume and pH were measured and flow rate calculated for each time period.

PRELIMINARY RESULTS

Control Data

Salivary Data. Sample collection and data analysis for the salivary data was completed. Salivary flow rate and pH appears to be relatively constant over the twelve hour sampling period using the Salivette system (Figure 1). Mean flow rate ranged between 50 to 60 ml/hr and mean salivary pH ranged between 6.5 and 7.0. No significant changes in flow rate or pH were noticed as a function of either the frequency of samples or time of day.

Performance Data. Data collection for this parameter is still in progress. Preliminary assessment of the data from the first few trials suggests that there is a subject dependent learning pattern with the program. In some individuals there was no improvement which suggests that these subjects did not exhibit a learning curve, but in others there was an apparent learning curve. There was a considerable intersubject variability in scores, but plateau scores were relatively constant for each individual. The learning curve appears to plateau after two test sessions with the math task and after five to eight test sessions with the memory task. There does not appear to be any difference between scores achieved in the morning versus scores achieved in the afternoon.

Scopolamine Study

Salivary Data. A limited number of subjects (intranasal n=8; intravenous n=7; oral n=8) have completed the study so far.

Preliminary results from these subjects were evaluated and reported here. Although all three dosage forms of scopolamine appear to decrease salivary flow rate and salivary pH, differences are apparent between the dosage forms (Figure 2). Intravenous scopolamine markedly inhibited salivary flow rate and lowered salivary pH. Flow was nearly totally inhibited in the first 15 minutes following administration of the drug, with a slow recovery to pre-dose flow in four to six hours. Salivary pH fell from the pre-dose mean of 6.6 to 5 in the first 90 minutes following intravenous administration of the drug, with slow recovery to pre-dose value by five hours. Intranasal scopolamine lowered salivary flow to about 5 ml/hr in about one hour with a slow recovery at nearly the same rate as with intravenous scopolamine. Salivary pH fell from the pre-dose mean of 6.7 to 4.8 in the first 90 minutes following the intranasal administration of the drug, which was a slightly larger decline than with the intravenous dosage form. Oral scopolamine caused a small decrease in flow in about one hour, with recovery to pre-dose flow by two hours. Salivary pH fell slightly from the pre-dose mean of 6.7 to 6.3 in the first three hours following oral administration.

Performance Data. Preliminary analysis of the performance parameters for the first few trials suggests that there is no change in task one or task two scores (percent correct, percent incorrect, percent timed out) following the administration of intranasal, intravenous, or oral scopolamine. None of the three dosage formulations adversely affected the mean correct response times for task one or task two. However, the mean correct response times for performance of task one following transition from task two and for performance of task two following transition from task one appear to be slightly longer in the first few hours following drug administration (Figure 3).

DISCUSSION

The flow rate of mixed saliva following the administration of anticholinergic drugs has been reported but was not well characterized. Grundhofer and Gibaldi found that oral administration of the anticholinergic drugs propantheline and hexocyclium lowered salivary flow rate by approximately 75% in two hours, with a gradual recovery to pre-dose rate by four to six hours (13,14). Brand *et al* reported that the administration of scopolamine at oral doses of 0.42 and 0.7 mg lowered salivary flow rate to approximately 50% of baseline by two to three hours (15). Gordon *et al* found that transdermal scopolamine lowered salivary flow rate by about 50% after 12 to 18 hours of transdermal administration (16). The present investigation is the first of its kind to utilize salivary characteristics for pharmacodynamic evaluation of dosage forms of scopolamine. Limited data collected so far indicate that there are differences in the pharmacodynamic effect as suggested by salivary flow rate after intranasal, intravenous, and oral administration of scopolamine, with a fast and pronounced effect from the

intravenous dosage form and a relatively slow and minimal effect from the oral dosage form. This appears to reflect the differences in pharmacokinetics and bioavailability of the three dosage forms.

Salivary pH depends on salivary flow (17). Therefore, it is anticipated that salivary pH decreased with the decreased flow rate. However, the lowest mean salivary pH following the intravenous and intranasal administration of scopolamine is lower than the 5.8 reported by Kreusser for unstimulated salivary glands. Salivary pH appears to be a good indicator of the differences in pharmacodynamic effect of different dosage forms, which may be a function of the pharmacokinetic properties of the three dosage forms.

Although psychometric testing has been used to assess the pharmacodynamic effects and side effects of centrally active drugs, limited information is available regarding the shapes and duration of learning curves or their interference in repeated measures testing. Schulz suggested that subjects should be given a period of familiarization with the testing procedures followed by repeated testing until the subjects achieve a predetermined coefficient of variation (18). Our results suggest that there is a learning curve with both tasks in our study, and that although there is some intersubject variability in individual plateaus for efficiency, the subjects require several tests to reach peak efficiency with the memory task. This data suggests that subjects require more experience with the program than is currently being provided prior to testing after administration of the drug.

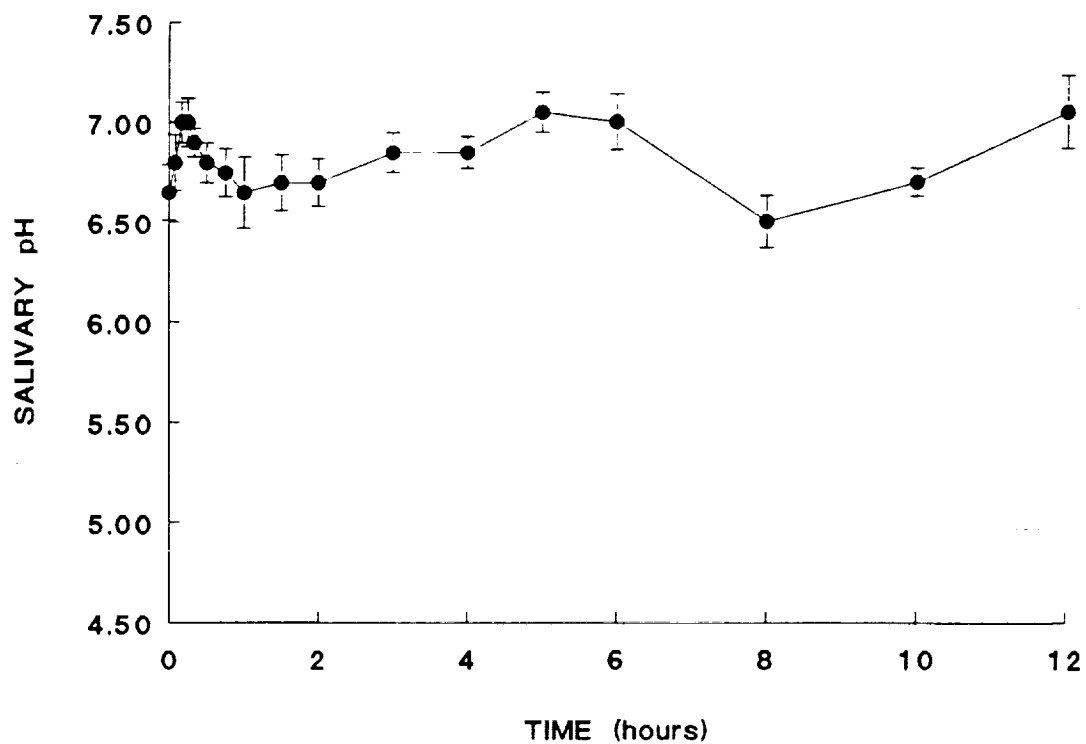
Traditional psychometric testing has documented that oral scopolamine impairs cognitive performance (6-12). Acetylcholine appears to be important for the processes of sustained attention and vigilance (7,19,20) as well as stimulus processing and memory storage (21-26). Although early studies indicated that scopolamine impaired longterm recall (23,27), more recent studies indicate that scopolamine impairs tasks involving sustained continuous attention or the storage of new information into memory rather than recall of information memorized before drug administration (8,11,12). In an investigation using a different investigational NTI performance battery, it was reported that scopolamine did not affect performance when using reaction time as a parameter for either spatial or verbal resources, but that scopolamine did impair the attention allocation system (28). Although preliminary, present data appear to support the findings that scopolamine impairs tasks requiring sustained continuous attention and processing of newly memorized information.

CONCLUSION

Limited information regarding the pharmacodynamics of scopolamine has been reported. This study was designed to quantitate the differences in pharmacodynamics between the intranasal,

intravenous, and oral dosage formulations of scopolamine. Data analysis for the completed studies will be performed and the relationship between the pharmacodynamics and the pharmacokinetics of the drug will be determined. It is anticipated that the results from these studies will provide new information that will be useful in the selection of appropriate drugs and dosage forms for the management of space motion sickness.

BASELINE SALIVARY pH



BASELINE SALIVARY FLOW

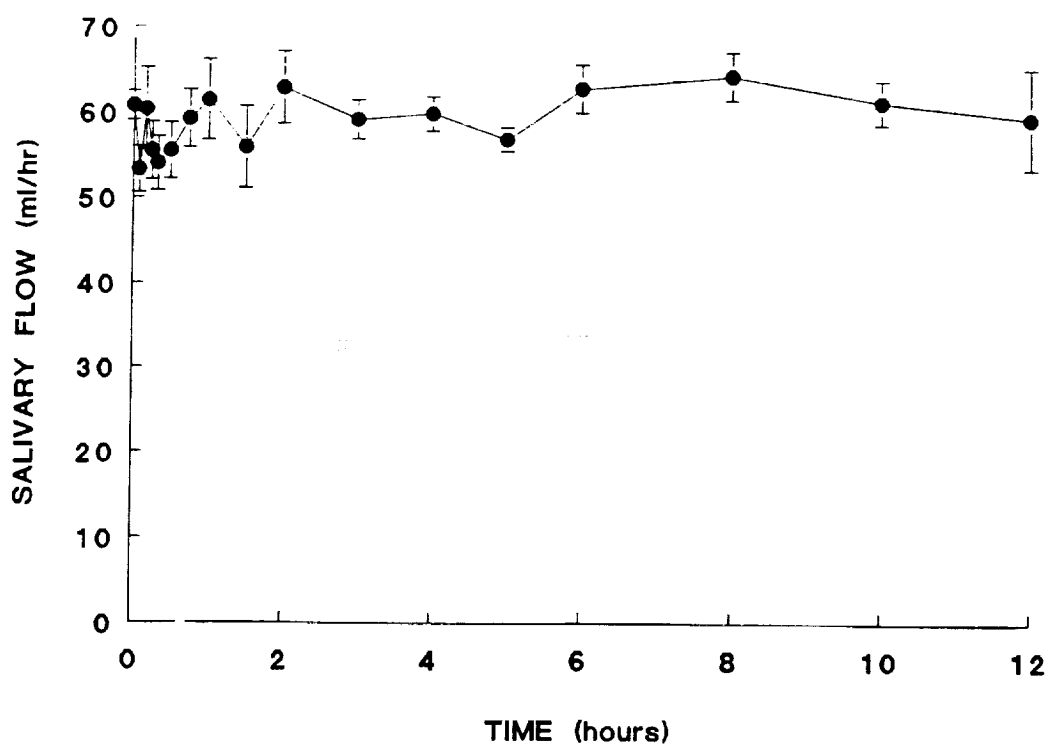
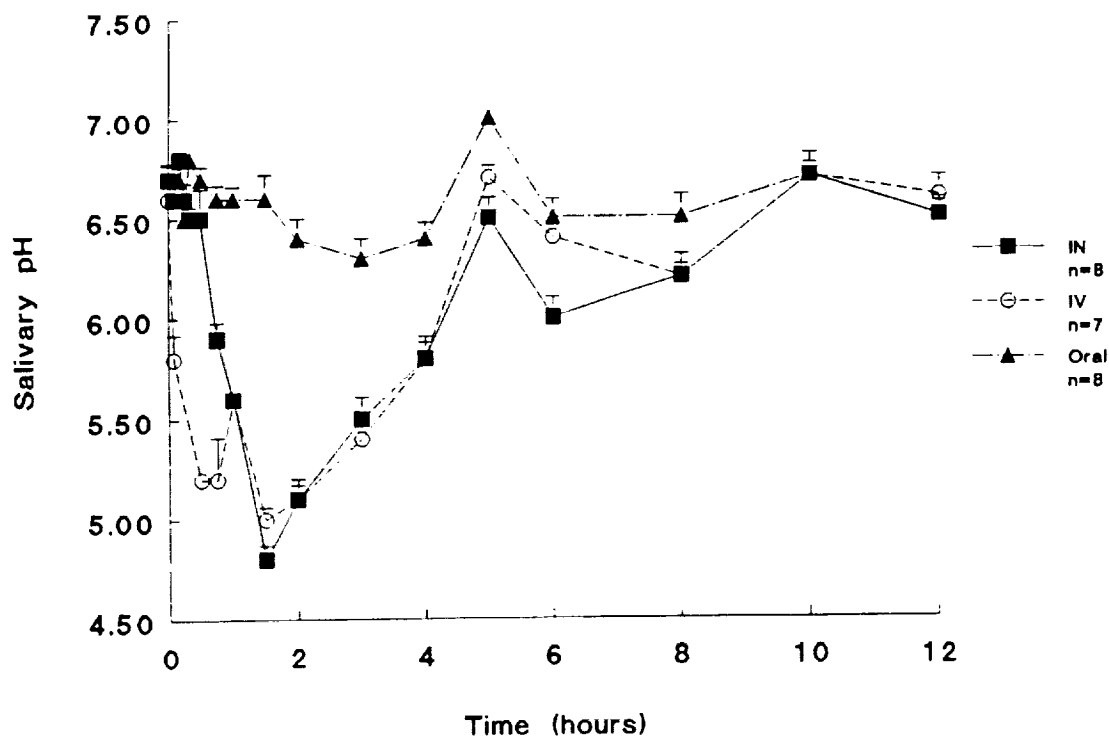


FIGURE 1.-BASELINE SALIVARY FLOW RATE AND pH.

Salivary pH



Salivary Flow

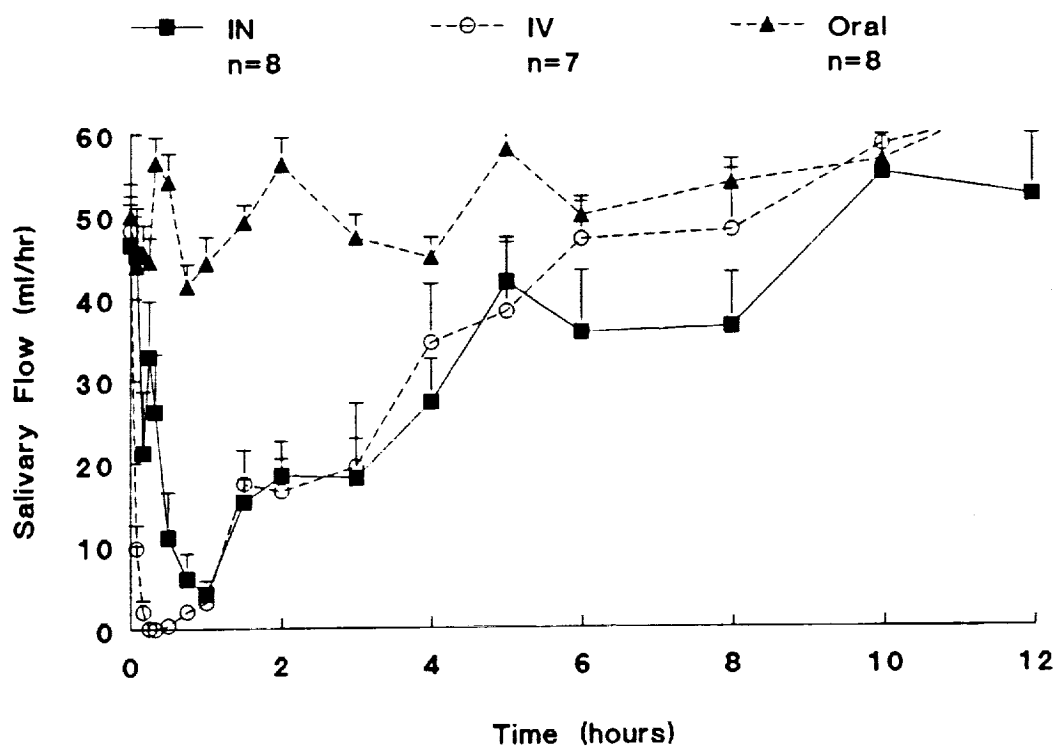
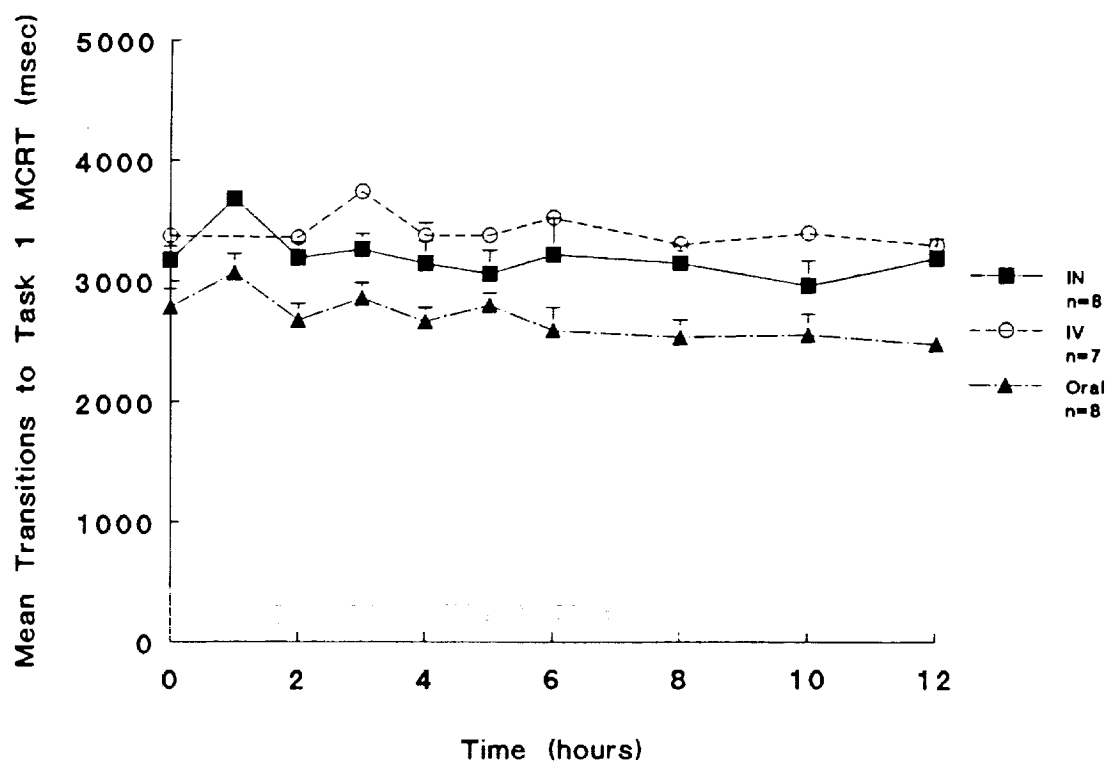


FIGURE 2.- SALIVARY FLOW RATE AND pH AFTER SCOPOLAMINE ADMINISTRATION.

Mean Transitions to Task 1 MCRT



Mean Transitions to Task 2 MCRT

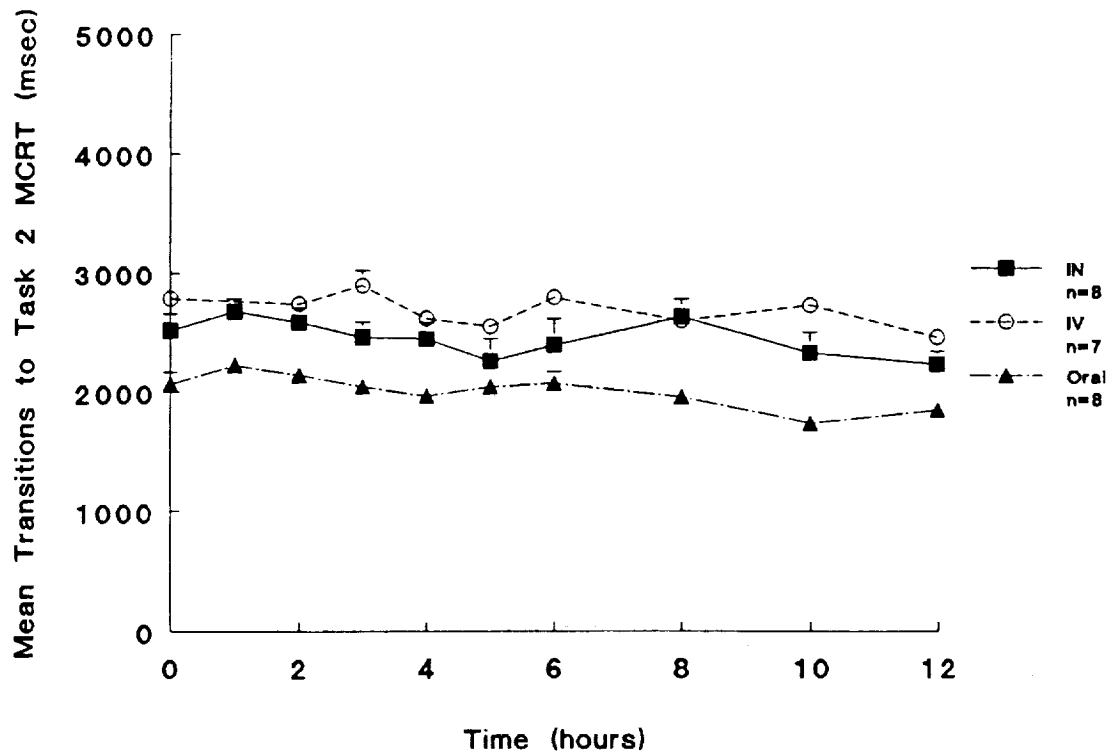


FIGURE 3.- MEAN CORRECT RESPONSE TIMES FOR TRANSITIONS TO TASK ONE AND TASK TWO AFTER SCOPOLAMINE ADMINISTRATION.

1. Homick JL, Vanderploeg JM. The Neurovestibular System. Space Physiology and Medicine, Nicogossian AE, Huntoon CL, Pool SL, eds, Lea & Febiger, Philadelphia, 2nd ed., 1989, p. 159.
2. Brand JJ, Perry WLM. Drugs used in motion sickness. Pharmacol Rev 1966; 18:895-24.
3. Wood CD, Manno JE, Manno BR, Redetzki HM, Wood M, Vekovius WA. Side effects of antimotion sickness drugs. Aviat Space Environ Med 1984; 58:113-16.
4. Kohl RL, Homick JL. Motion sickness: a modulatory role for the central cholinergic nervous system. Neurosci Biobehav Rev 1983; 7:73-85.
5. Wood CD, Graybiel A. Theory of antimotion sickness drug mechanisms. Aerospace Med 1972; 43:249-52.
6. Schmedtje JF Jr., Oman CM, Letz R, Baker EL. Effects of scopolamine and dextroamphetamine on human performance. Aviat Space Environ Med 1988; 59:407-10.
7. Poulton EC, Edwards RS. Interactions, range effects, and comparisons between tasks in experiments measuring performance with pairs of stresses: mild heat and 1 mg of L-hyoscine hydrobromide. Aerospace Med 1974; 45:75-41.
8. Parrott AC. The effects of transdermal scopolamine and four dose levels of oral scopolamine (0.15, 0.3, 0.5, and 1.2 mg) upon psychological performance. Psychopharmacology 1986; 89:347-54.
9. Parrott AC. Transdermal scopolamine: effects of single and repeated patches upon psychological task performance. Neuropsychobiology, 1987; 17:53-9.
10. Rusted JM, Warburton DM. The effects of scopolamine on working memory in healthy young volunteers. Psychopharmacology 1988; 96:145-52.
11. Rusted JM. Dissociative effects of scopolamine on working memory in healthy young volunteers. Psychopharmacology 1988; 96:487-92.
12. Flicker C, Serby M, Ferris SH. Scopolamine effects on memory, language, visuospatial praxis and psychomotor speed. Psychopharmacology 1990; 100:243-50.
13. Grundhofer B, Gibaldi M. Biopharmaceutic factors that influence effects of anticholinergic drugs: comparison of propantheline, hexocyclium, and isopropamide. J Pharm Sci 1977; 66:1433-5.

14. Gibaldi M, Grundhofer B. Biopharmaceutic influences on the anticholinergic effects of propantheline. Clin Pharmacol Ther 1975; 17:457-61.
15. Brand JJ, Colquhoun WP, Perry WLM. Side effects of L-hyoscine and cyclizine studied by objective tests. Aerospace Med 1968; 39:999-1002.
16. Gordon C, Ben-Aryeh H, Attias J, Szargel R, Gutman D. Effect of transdermal scopolamine on salivation. J Clin Pharmacol 1985; 25:407-12.
17. Kreusser W, Heidland A, Hennemann H, Wigand ME, Knauf H. Mono- and divalent electrolyte patterns, pCO₂ and pH in relation to flow rate in normal human parotid saliva. Eur J Clin Invest 1972; 2:398-406.
18. Schulz R, Reimann IW. Practice effect of volunteers in repeated psychometric testing. How to handle this intervening variable in clinical pharmacology studies? Meth and Find Exptl Clin Pharmacol 1988; 10:657-661.
19. Colquhoun WP. Effects of hyoscine and meclizine on vigilance and short-term memory. Brit J Industr Med 1962; 19:287-96.
20. Wesnes K, Warburton DM. Effects of scopolamine and nicotine on human rapid information processing performance. Psychopharmacology 1984; 82:147-50.
21. Crow TJ, Grove-White IG. An analysis of the learning deficit following hyoscine administration to man. Br J Pharmac 1973; 49:322-7.
22. Drachman DA. Memory and cognitive function in man: does the cholinergic system have a specific role? Neurology 1977; 27:783-90.
23. Ghoneim MM, Mewaldt SP. Effects of diazepam and scopolamine on storage, retrieval and organizational processes in memory. Psychopharmacologia 1975; 44:257-62.
24. Jones DM, Jones MEH, Lewis MJ, Spriggs TLB. Drugs and human memory: effects of low doses of nitrazepam and hyoscine on retention. Br J Clin Pharmac 1979; 7:479-83.
25. Deutsch JA. The cholinergic synapse and the site of memory. Science 1971; 174: 788-94.
26. Warburton DM, Wesnes K. Drugs as research tools in psychology: cholinergic drugs and information processing. Neuropsychobiology 1984; 11:121-32.

27. Peterson RC. Scopolamine induced learning failures in man. Psychopharmacology 1977; 52:283-9.
28. Botha, CJ: The effects of low-dose scopolamine on selective cognitive functions. M.S. Thesis, Wright State University, 1990.

512-74
N91-274157
P-13

HIGH ACCURACY OPTICAL RATE SENSOR

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1990

Johnson Space Center

S 2864095

Prepared By:	J. Uhde-Lacovara, Ph.D.
Academic Rank:	Assistant Professor
University & Department	Stevens Institute of Technology Dept. of Electrical Engineering and Computer Science
NASA/JSC	
Directorate:	Engineering
Division:	Navigation, Control and Aeronautics
Branch:	Navigation and Guidance Systems
JSC Colleague	Indulis Saulietis
Date Submitted	August 17, 1990
Contract Number	NGT-44-005-803

ABSTRACT

Optical rate sensors, in particular CCD arrays, will be used on Space Station Freedom to track stars in order to provide inertial attitude reference. An algorithm to provide attitude rate information by directly manipulating the sensor pixel intensity output is presented. The star image produced by a sensor in the laboratory is modeled. Simulated, moving star images are generated, and the algorithm is applied to this data for a star moving at a constant rate. The algorithm produces accurate derived rate for the above data. A step rate change requires two frames for the output of the algorithm to accurately reflect the new rate. When zero mean Gaussian noise with a standard deviation of 5 is added to the simulated data of a star image moving at a constant rate, the algorithm derives the rate with an error of 1.9 per cent at a rate of 1.28 pixels per frame.

INTRODUCTION

Optical rate sensors will be of great use on Space Station Freedom as part of the onboard guidance, navigation, and control system. These sensors can be used to track stars to provide inertial attitude reference. The information may also be used by astronomical experiments to allow reference to a known star catalog to provide precise pointing of instruments. Optical sensors can be used on Freedom to keep track of other nearby objects such as incoming orbiters. Co-orbiting platforms may also use optical sensors for their GN&C needs.

Because optical sensors will provide attitude information to Space Station, it would be useful to derive attitude rate could from these sensors. An optical rate sensor is currently being developed in the Navigation, Control, and Aeronautics Division at Johnson Space Center. The sensor will look at stars to obtain both attitude information and to derive attitude rate. A Videk Megapixel camera is supported by a Compaq Deskpro 386 for data capture and processing. The Videk camera uses a Kodak charge coupled device (CCD) array. This array is made up of 1320 horizontal by 1035 vertical pixels. The pixels are 6.8 microns square and the array is full fill. The camera is capable of 7 still frames per second; it has a grey scale of 256 levels.

In previous work on the rate sensor [1,2], two different approaches were taken. The first approach was to differentiate the centroid of the star image to obtain derived rate. A discrete-time, nonrecursive, filtering algorithm was used. The Space Station accuracy and bandwidth requirements could not simultaneously be met using this technique. The second approach was to directly manipulate the image sensor data to obtain derived rate. This technique was successfully simulated for a highly idealized star image in the presence of additive Gaussian noise.

Source stars were simulated in the laboratory by a fiber-optic strand illuminated by a light source. This laboratory setup is shown in Figure 1. The fiber-optic strand is mounted in a metal plate for stability. A filter plate with various apertures and colored filters is placed in front of the fiber optic. This assembly is attached to a micrometer drive to allow precise control of the movement of the optical fiber. The simulated star light was then passed through a collimator to make the light appear to emanate from a point source at infinity. After the light passes through the collimator, it reaches the camera.

The camera data is captured using a frame grabber.

STAR IMAGE MODELING

A three dimensional plot of a star image obtained from the above laboratory experiment is shown in Figure 2. The integration time, that is the time for which the camera shutter was open and the array was exposed to the star light, was 100 milliseconds. The data was captured for an array of 68 pixels in the x-direction by 32 pixels in the y-direction. The plot presented is actually for three frames of the same star image averaged for noise reduction.

The problem was first reduced from a two-dimensional problem to a one-dimensional problem. Column sums were performed on the array to produce the x-direction pixel sums, and row sums were performed to produce the y-direction pixel sums. This gave a one-dimensional image which was fit by a Gaussian curve of the form;

$$I(x) = be^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The original star image was thresholded to remove noise generated in the unilluminated pixels. All pixels with intensities less than or equal to 3.0 had their intensities set equal to zero for the thresholded star. This produced an image in which all pixels with non-zero intensity were contiguous. The center of mass and moment of inertia of the x-direction column sums were found for the thresholded star image as follows:

$$\mu = 31.1 \text{ pixels}$$

$$\sigma = 3.03 \text{ pixels}$$

Figure 3 shows a plot of the x-direction pixel sums versus a generated Gaussian with the same μ and σ , and with a peak intensity of 1,1533.

In general, a moving star image can be simulated by generating a set of Gaussian curves in which the center of mass of the image changes as a function of time. The equation for the intensity seen by a given pixel at a position x, and time, t is given below:

$$I(x,t) = be^{-\frac{(x-\mu(t))^2}{2\sigma^2}}$$

A set of Gaussian curves was generated for a star image with a constant rate of motion in the x-direction. A Gaussian with a peak intensity of 1200, and moment of inertia of 2.5 pixels was used to attempt to fit the central portion of the thresholded x-direction column sums. A rate of 1.28 pixels per frame for 30 frames was used for the generated images to simulate data obtained in the laboratory.

RATE SENSOR ALGORITHM

Derivation of Algorithm

Centroiding a star image to derive rate is essentially a curve-fitting problem in which the peak intensity and spread (moment of inertia) of the image are unknown. Stars of different magnitudes will have different peak intensities for the same integration time. If different integration times are used for the same star, the spread of the image may change. Knowledge of the Gaussian function, and the type of rate to be sensed is exploited to eliminate the need to determine the above parameters directly.

The algorithm is derived for the one-dimensional case, and is then extended to two-dimensions. It is important to remember that the star data to be processed is discrete in time, space, and amplitude. The algorithm is first derived by assuming the rate is constant. This implies that the position of the center of mass is the product of the rate in pixels per frame times the frame number plus an offset to account for the original position at the first frame. That is

$$\mu(f) = rf + \mu_0$$

If the offset is made zero, the intensity for a given pixel, n , at a given frame, f , for the constant rate is shown below:

$$I(n,f) = be^{-\frac{(n-rf)^2}{2\sigma^2}}$$

Taking the ratio of the intensity for a given pixel, n , for two successive frames gives

$$\begin{aligned} \left[\frac{I(n, f+1)}{I(n, f)} \right] &= \frac{be^{-(n-r(f+1))^2/2\sigma^2}}{be^{-(n-rf)^2/2\sigma^2}} \\ &= e^{-(r^2 + 2r^2f - 2nr)/2\sigma^2} \end{aligned}$$

The natural logarithm of the ratio yields

$$\ln \left[\frac{I(n, f+1)}{I(n, f)} \right] = (2nr - r^2(2f + 1))/2\sigma^2$$

This is a linear function of f with slope proportional to r^2 and σ^2 .

Similarly, the natural logarithm of the intensities for a given frame, f , for two adjacent pixels is given by

$$\ln \left[\frac{I(n+1, f)}{I(n, f)} \right] = -(2n + 1 - 2rf)/2\sigma^2$$

This is also a linear function of f with slope proportional to r and σ^2 .

Taking the negative of the slope found previously, and dividing by the slope found above, yields the rate, r . The logarithm of the ratio is equivalent to subtracting the logarithms of the intensities. It should be noted that the peak intensity, b , and the spread of the image, σ , need never be found.

Rate Algorithm

- 1) Find the natural logarithms of the pixel intensities
- 2) Find the frame to frame differences of the log intensities for a given pixel; do this for several pixels and pairs of frames.
- 3) Find the frame to frame differences of the above; let the result of these

- manipulations be represented by the constant, $K_f = -r^2/\sigma^2$.
- 4) Find the pixel to pixel differences of the log intensities for a given frame; do this for several frames and pairs of pixels.
 - 5) Find the frame to frame differences of the above; let the result of these manipulations be represented by the constant, $K_p = r/\sigma^2$.
 - 6) Divide $-K_f$ from 3) by K_p from 5) to obtain the rate.

RESULTS

Constant Rate Without Noise

The algorithm was applied to the simulated moving star data previously described which had a rate of 1.28 pixels per frame for 30 frames. The resulting derived values are as follows:

$$\begin{aligned} K_f &= -0.2621; \\ K_p &= 0.02048; \\ r_{\text{derived}} &= 1.2800; \text{ (carried out to machine precision)} \end{aligned}$$

Step Rate Change Without Noise

Fifteen frames of data were simulated as before with a rate $r_1=1.2800$ pixels per frame. At the sixteenth frame, the rate was increased by ten per cent to $r_2=1.4080$. Fifteen additional frame were generated at the new rate for a total of 30 frames. The constant K_f settled to its new value by the eighteenth frame; the constant K_p settled to its new value by the seventeenth frame. The correct new rate could, therefore, be derived by the eighteenth frame. Frames 16 and 17 produced outputs which could not be used to correctly derive the rate. This region of indeterminacy will be dealt with in the same way as noisy images. This method will be described in the next section.

Constant Rate Change With Noise

The algorithm was tested in the presence of noise. Gaussian noise with zero mean and standard deviation of 5 was generated. The absolute value of the noise

was added to the 30 frames of generated star images moving at a constant rate of 1.2800 pixels per frame. Figure 4 shows a log plot of the generated data. As seen from Figure 4, the low intensity values were significantly distorted by the noise. Thresholding was performed on the data; all log intensities less than 4.5 were replaced with 0. The first five steps of the algorithm were performed on the noisy, thresholded data. Figure 5 shows a plot of the output of the algorithm at step 3. The spikes and numerous zero values which occur are artifacts of the thresholding. The desired data is the horizontal line just below the zero line. This data needs to be extracted from the output. A similar plot is seen for the output of step 5.

A histogramming technique similar to that employed in previous work was applied to this data [2]. Histograms of the frame differences and of the pixel differences were made. The peak frequency of zero was excluded because this is an artifact of the thresholding. Taking the second highest peak frequencies, the following values are obtained:

$$\begin{aligned} K_f &= -0.2531; \\ K_p &= 0.1940; \\ r_{\text{derived}} &= 1.3044; \end{aligned}$$

This represents an error of 1.9 per cent in the derived rate.

Improvements in this error performance might be obtained by increased thresholding, preprocessing to reduce the noise in the image and more sophisticated histogramming techniques. It should also be noted that histogramming techniques will handle a step rate change; the false rates obtained during the transition phase will have a low probability of occurrence, and will therefore not be included in the rate calculations. The rate is constantly updated by removing stale estimates as fresh data points are added to the histogram.

Laboratory Data

Figure 6 shows the total intensity versus the column sums for the x-direction pixels. This is 30 frames of data for a star moving at approximately four thirds of a pixel per frame. It can be seen from this plot that there are noticeable dips in the intensity at regularly spaced intervals of 16 pixels. This phenomenon was noticed toward the end of the summer program. An explanation of it was not obtained in time to compensate fully for the effect in

applying the algorithm. It may be stated that, for this corrupted data, an approximation to the rate may be obtained from applying the algorithm, but the results are inaccurate at this point. The sixteen pixel periodicity in the sensor response is related to the physical process used to manufacture the CCD array.

CONCLUSIONS

The rate algorithm described works well for simulated data for the optical rate sensor. The open issues which will be addressed immediately are those of refining the model of the star image which the sensor produces, and compensating for the systematic intensity errors introduced by the CCD array. A parametric study of the algorithm, in conjunction with real sensors, needs to be performed in order to determine such quantities as allowable noise, frame rate, pixel size, number of grey levels, number of pixels and frames to be processed at one time, etc.

The algorithm is easily extended to two-dimensions by taking both the x-direction column sums to obtain the x-direction rate, and the y-direction row sums to obtain the y-direction rate in field coordinates, the field coordinates can then be transformed to the body coordinates of the vehicle.

REFERENCES

- [1] Uhde-Lacovara, J., "Analysis of the Continuous Stellar Tracking Attitude Reference (CSTAR) Attitude Rate Processor", NASA/ASEE Summer Faculty Fellowship Program, 1986, Volume 2 (N87-25884 19-85).
- [2] Uhde-Lacovara, J., "Optical Rate Sensor Algorithms", NASA/ASEE Summer Faculty Fellowship Program, 1989, Volume 2 (N90-24985 18-80).

FIGURE 1

Rate Sensor Optical Bench

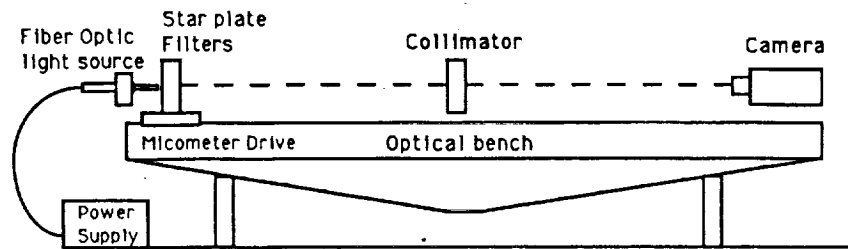


FIGURE 2

STAR IMAGE: 100 msec integration time

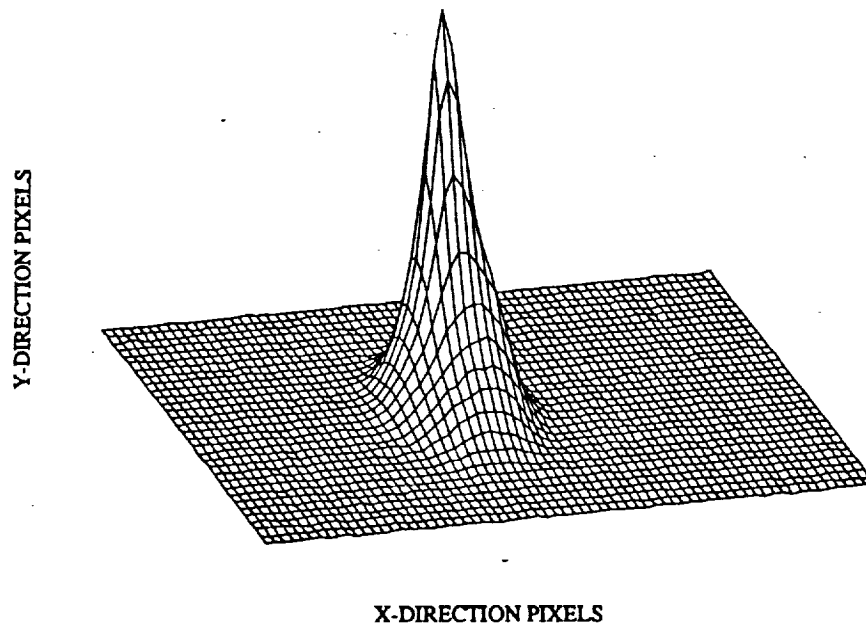


FIGURE 3

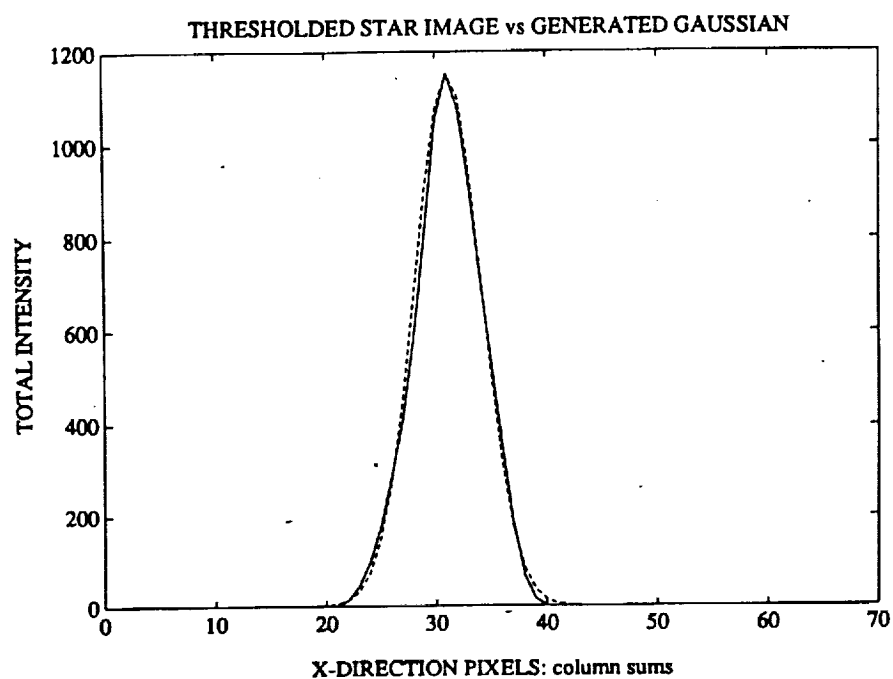


FIGURE 4

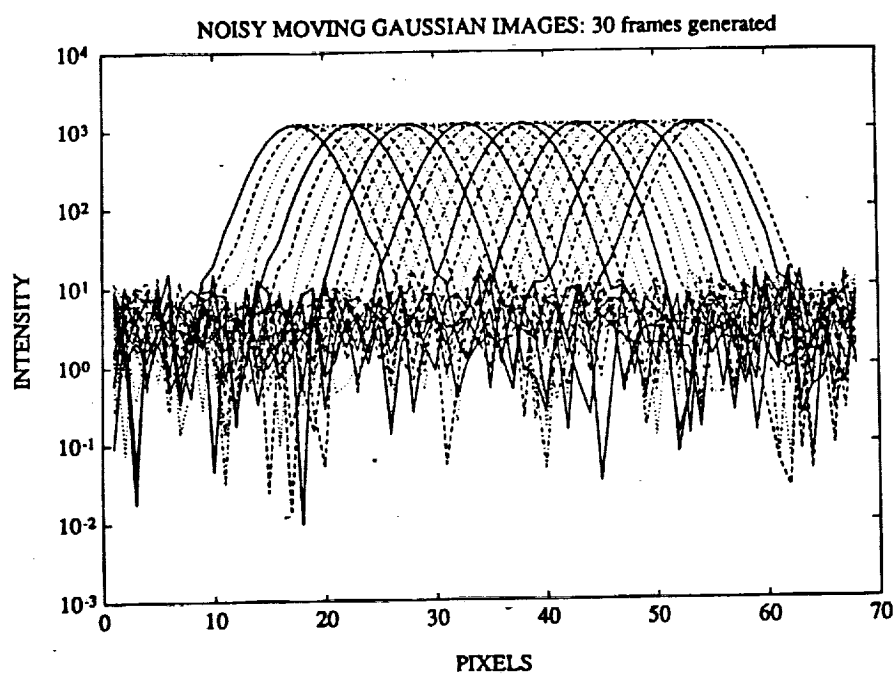


FIGURE 5

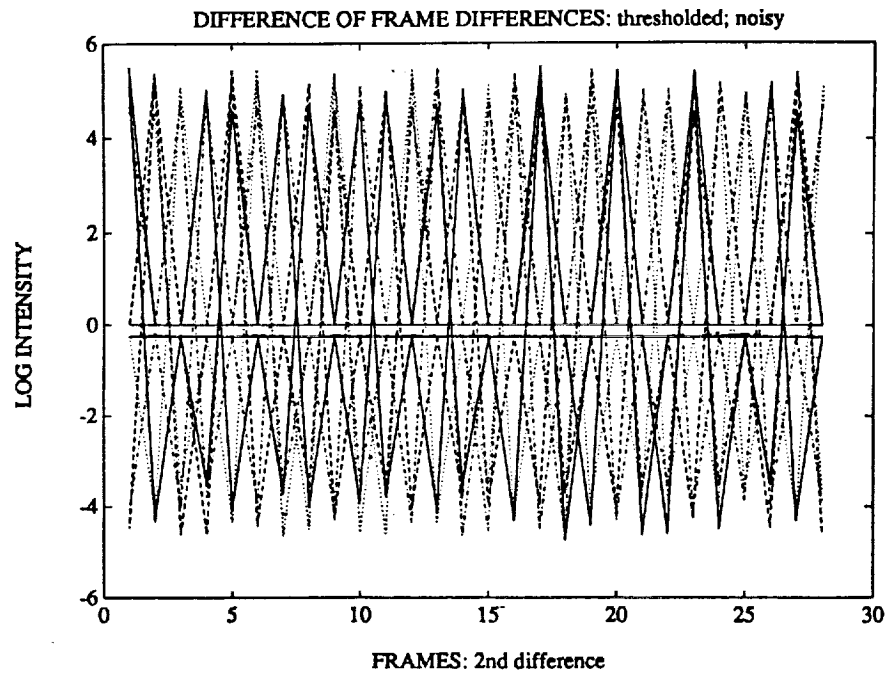
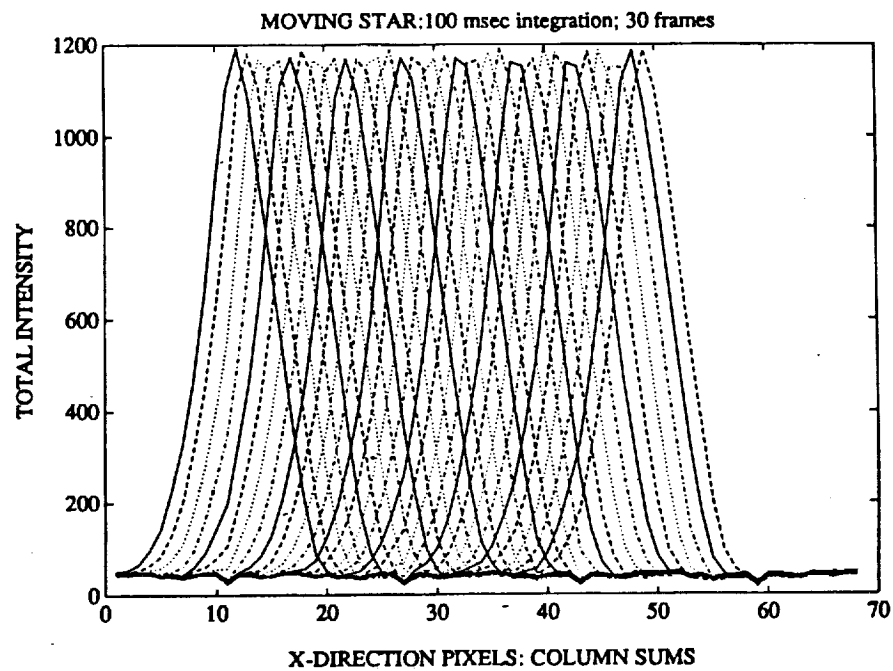


FIGURE 6



5/3-83
N91-27116²⁰²²⁸
P-6

PLANETARY SPACECRAFT COST MODELING UTILIZING
LABOR ESTIMATING RELATIONSHIPS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

M3452869

Prepared by:	Raymond Williams, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Mississippi Valley State University Computer Science and Mathematics Department Itta Bena, Mississippi 38941
NASA/JSC	
Directorate:	New Initiatives Office
Division:	Project Planning & External Affairs
JSC Colleague:	Jared R. Woodfill
Date Submitted:	August 10, 1990
Contract:	NGT 44-005-803

ABSTRACT

A basic computerized methodology is presented for estimating labor hours and cost of unmanned planetary and lunar programs. The user friendly methodology designated Labor Estimating Relationship/Cost Estimating Relationship (LERCER) organizes the forecasting process according to vehicle subsystem levels. The level of input variables required by the model in predicting cost is consistent with pre-Phase A type mission analysis. Twenty-one program categories were used in the modeling: sixteen spacecraft subsystem categories (structures and devices, thermal control, cabling and pyrotechnics, propulsion, attitudes and articulation control, telecommunication, antennas, command and data handling, RTG, solar batteries, aerodeceleration, radar, line-scan, vidicon, particle and field, remote sensing, and direct sensing and sampling); and five support categories (program management, system support and ground equipment, launch +30 days operation and ground software, image data development, and science data development). To develop this model, numerous LER and CER studies were surveyed and modified where required. The most recent models that estimated cost of historical data to within 20% of the actual cost were selected. The result of the research along with components of the LERCER program are reported. On the basis of the analysis the following conclusions were established: (1) spacecraft cost is highly correlated with man (labor) hours; (2) no user friendly program was found that linked LERs with CERs to produce spacecraft cost as computer output; (3) the cost of a spacecraft is best estimated by summing cost of individual subsystems; (4) the group accumulation labor estimation method proved to be a useful method for finding total manpower labor hours and cost of a spacecraft; (5) the cost of 1 kg mass across subsystems can vary from \$152K (structure and devices) to \$6.6M (thermal control, cabling and pyrotechnics); (6) program labor cost constitutes approximately 30 percent of the total program cost.

INTRODUCTION

Space program cost estimation in pre-Phase A stage of development remains a challenge for cost forecasters. A central idea presented in this study is that cost forecasting can be improved by selecting manhours as the primary cost unit. One major advantage of manhours as the basic cost unit over forecasting total program dollar amount is the separation of inflationary factors. Two programs costed at different times can only be compared if some inflationary factor is applied to the older program. The ideal inflation factors are often difficult to formulate for total program cost and often fail to be good representative of the actual financial conditions within industry. Industry still has not been able to mass reproduce spacecraft and consequently the total cost of each subsystem is not significantly decreased through additional production.

The purpose of this study was (1) to model spacecraft cost as a function of manhours; (2) to interconnect manhour equations and cost equations; (3) to develop a computer program that uses LER output as input into CERs producing the cost of a spacecraft as final output; and (4) to extract from the literature those labor estimating relations (LERs) that determine the total cost of a spacecraft by summing the cost of its individual subsystems.

LABOR ESTIMATING RELATIONSHIPS (LERs)

Early and adequate estimation of program cost requirements should be of major concern to both the government and government contractors. Experience indicates that final cost of advanced technology activities has been extremely higher than that originally predicted. Somehow, prediction procedures have resulted in large errors in costing spacecraft programs. Thus, there is a rapidly growing need for a credible spacecraft cost model. It should be noted that a cost model does not necessarily represent factual information, but rather it is based on experimental approximations. Costing spacecraft by using dollar amounts as the basic cost unit is the method reported by most cost forecasting studies. This paper uses, instead of dollar expenditure, manhours (labor hours) as the basic cost unit. Forecasting labor hours has several advantages over directly forecasting total program dollar amount. One such advantage,

and most important, is that manhours separate inflationary factors from the model. Inflationary factors are very difficult to formulate for total program cost and often time are not consistent with inflation rates within industry. Additional information reported by industry reveals that space industry is still unable to mass produce spacecraft. That is, the cost to make additional copies of the same flight unit is not significantly reduced from that of the original unit. Thus, it appears that project hardware cost is directly related to the manhours involved in the development, fabrication and testing. Hence, the manpower approach to cost forecasting should provide management with a tool to evaluate the cost of new projects in light of formulated manpower estimating relationships.

COST ESTIMATING RELATIONSHIPS (CERs)

Historically, early estimates of spacecraft project cost have been considerably less than the final actual cost. Spacecraft have been very expensive to produce because of stringent weight and performance requirements, heavy emphasis on reliability, and small production quantities. Various parametric cost estimating models have been developed from experience of the past 30 years, and those models reproduce the cost of the traditional spacecraft with acceptable accuracy. Due to the limited data base of manned space vehicles, a spacecraft cost estimating procedure based on historical data often contains errors. Mathematical relationship errors will exist in any estimating procedure which attempts simplification of actual cause and effect by empirical approximation. A good cost model, of course, is one that strives to minimize these inevitable errors. Cost regression equations for spacecraft subsystems are typical of the type

$$Y = AX^b \text{ or } y = A + BX^c$$

Where Y = cost and X = weight or some other subsystem characteristic. Those cost models that were shown to predict the cost of historical vehicles to within 20 percent of the actual cost were selected or developed for this study.

LERCER PROGRAM APPLICATION/OPERATION

All spacecraft information for use with the LERCER computer program is found within the LERCER folder on diskette. To operate the program open the LERCER folder and all files therein. Input data description is found at the top of each column. Row 5 of column A gives instructions for handling spacecraft flight units.

The LERCER program strives to use features resident in EXCEL and the Macintosh system in a user friendly fashion. Among these features is the "GOTO" instruction. Selecting the "GOTO" from the "FORMULA" column on the menu bar opens a listing of all 21 subsystems and other pertinent categories listed in alphabetical order. Scroll through this list and make your selection. Control will then be transferred to this selection in the program. You may at this point make modifications in the input data if you so desire.

Each file with the surname TOTAL represents an accumulation of all spacecraft calculations for the total of a particular spacecraft program. Sample calculations for two planetary programs are included in LERCER: one of which is the Mars Rover Sample Return (MRSR) program developed by Martin Marietta Corporation and the second calculation is for the Mass Matrix (MM) program developed in-house at JSC.

The LERCER program has a fundamental feature: to use pre-Phase A weight estimates for each subsystem to estimate manpower hours; to input the forecasted manhours into a corresponding cost equation for forecasting cost of planetary subsystems; and finally, to calculate the total forecasted cost.

The LERCER program features facilitate this fundamental operation. Initial input data is entered in ledger fashion in a single column of the EXCEL spreadsheet matrix. The input data consists of listed constants selected by the user, adjustment factors, and the subsystem mass. Each LERCER equation is listed adjacent to the entry data column to assist the user. Again, a designated ledger column serves for listing of all LERs and CERs.

Total cost includes two categories: non-recurring (design, development, test, and evaluation (DDTE)) cost and recurring flight hardware (FH) cost. Each subsystem has two cost forecasting equations listed on separate LERCER lines. The total cost for each spacecraft is found by summing the total cost of individual subsystems. Similarly, total recurring cost of each spacecraft is found by summing the recurring cost of individual subsystems. Non-recurring cost may be obtained by subtracting recurring cost from the total cost.

SUMMARY AND CONCLUSION

Early estimates of planetary spacecraft programs have been significantly less than the final actual cost. The ratio of actual cost to forecasted cost has been reported as high as 4:1. Manhour expenditures as a major costing unit was found to substantially lower this ratio. The manhours needed to build a planetary spacecraft may be adequately estimated and costed for initial pre-Phase A design utilizing credible historical data and regression analysis within a programmed methodology. The LERCER is such a system developed by this study. This program was developed for the New Initiatives Office (NIO) at JSC. Content of the methodology includes a slightly modified version of Science Applications Incorporated (SAI) manhour model and cost estimating relationships developed by the study. Dry mass is expressed in kilograms and cost in 1990 million-dollar units.

514-39
N91-271170229

STRUCTURAL CONCEPT STUDIES FOR
A HORIZONTAL CYLINDRICAL LUNAR HABITAT
AND
A LUNAR GUYED TOWER

P-15

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1990

Johnson Space Center

CS 204368

Prepared by:	Paul K. Yin, Ph.D., P.E.
Academic Rank:	Associate Professor
University and Department:	U.S. Coast Guard Academy Engineering Department New London, Conn. 06320
NASA/JSC	
Directorate:	Engineering
Division:	Systems Engineering
Branch:	Systems Definition
JSC Colleague:	Clarence Bell, Ph.D.
Date Submitted:	August 10, 1990
Contract Number:	NGT-44-005-803

ABSTRACT

A conceptual structural design of a horizontal cylindrical lunar habitat is presented. The design includes the interior floor framing, the exterior support structure, the foundation mat, and the radiation shielding. Particular attention has been given on its efficiency in shipping and field erection, and on selection of structural materials.

Presented also is a conceptual design of a 2000-foot lunar guyed tower. A special field erection scheme is implemented in the design. In order to analyze the over-all column buckling of the mast, where its axial compression includes its own body weight, a simple numerical procedure is formulated in a form ready for coding in FORTRAN. Selection of structural materials, effect of temperature variations, dynamic response of the tower to moonquake, and guy anchoring system are discussed.

Proposed field erection concepts for the habitat and for the guyed tower are described.

INTRODUCTION

The Systems Engineering Division of the NASA/JSC is studying the development of permanent human presence on the moon. The development includes a relatively large habitat to provide long term human habitation, antenna towers for communication, among others.

The author designed the internal framing and the external foundation for a spherical inflatable lunar habitat during the summer of 1989 (A concept developed within NASA/JSC) [1,2]. With the experience gained on the first design, the Systems Engineering Division is considering a horizontal cylindrical habitat in addition to a spherical one. A conceptual structural design of this second concept (excluding the inflatable air enclosure) is presented in this report. The structure was analyzed by Ed Robertson of Systems Engineering Division, Systems Definition Branch using SDRC's IDEAS program [3]. An order-of-magnitude estimate of total structural dead weight is presented. The estimate is based on the structural members of Aluminum Lithium 8090-T8771 for the inside of the habitat and Magnesium alloy ZCM 711 [4] for the outside of the habitat, all within the AISC stress limitations [6] under the specified design live loads. The architectural and functional requirements including design live loads were provided by K. J. Kennedy of the Systems Engineering Division. The interior and exterior structural framings are to be pre-assembled, packaged for shipping, and automatically deployed into its final installed configuration with minimum effort required for

field assemblage.

The development of permanent human presence on the moon also includes systems which may require installations of tower structures. Such systems may be for the lunar surface communication, power transmission, and area illumination. Given the radius of the lunar sphere the height of the towers is dictated by the distance between two adjacent towers. The longer the distance between towers, the higher the towers must be. While the communication area covered by a tower increases with the square of the tower's height the structural dead weight increases, generally speaking, with the cubic of the tower's height. This implies that the optimum height of the tower for lunar surface communication is zero. Further trade-off study on the use of towers for communication to the far side of the moon is needed including other factors and other possible alternatives.

For a surface area on the moon covered by a 30-mile radius (a limitation on an un-pressurized lunar rover) the height of the tower is approximately 2000 feet. A conceptual design of a 2000-foot lunar guyed tower, its materials, and field erection scheme are presented.

The installation of a 2000-foot guyed tower on the surface of the moon represents some unique design considerations. These considerations were studied and are discussed in this report.

THE HABITAT

General Concept

The basic concept of the horizontal cylindrical lunar habitat is similar to that of the spherical inflatable lunar habitat studied in the summer of 1989 [1,2] except that the over-all shape of the inflatable air enclosure is horizontal cylindrical, the regolith shielding is self supporting, the site preparation involves the removal of only the first one-foot layer of the lunar surface soil, and the framing is more efficient for shipping and field erection (Figure 1).

Design Requirements

The requirements, objectives and constraints are:

1. A minimized total structural dead weight.
2. Satisfaction of the Specification for the Design, Fabrication, and Erection of Structural Steel for Buildings (AISC) [6].
3. Efficient for shipping and field erection.

4. A design live load of 50 earth pound per square foot (2.40 kPa) for the upper floor and 100 earth pound per square foot (4.79 kPa) for the lower floor (from K. J. Kennedy, Systems Engineering Division, JSC).
5. Architectural requirements as given by K.J. Kennedy of the Systems Engineering Division (Figure 1).

Interior Structural Framing

A highly efficient flat space truss (a pallet, developed by the Structural Engineering Division of JSC [12]) was determined to be suitable for both the two interior floors. The truss can collapse into a highly compact bundle for shipping and be automatically deployed into its final configuration at the site. The interior floor trusses for both the two levels (Figure 1) are about one-foot deep and are composed of 0.394-inch (1-cm) diameter round bars of Aluminum Lithium 8090-T8771 all of equal length for the upper and the lower cord planes and the 0.315-inch (0.8-cm) web members. Each floor truss is supported at ten points through which the area live loading is transferred across the fabric to the exterior support framing (Figure 1).

Exterior Support Structure

The support structure outside the habitat is designed so that the structure can be retracted into a bundle for shipping with its length slightly greater than the diameter of the cylindrical air enclosure (Figure 1).

The interior floor trusses and the exterior support structure were analyzed by Ed Robertson of the Systems Engineering Division, JSC using SDRC's IDEAS [3]. Under the design live loads, the maximum stresses and deflections were all found within AISC's limitations [6].

Foundation

The foundation mat provides enough bearing area for a factor of safety of 4 against ultimate bearing capacity of the lunar soil at about one-foot depth based on the lunar soil properties given by Mitchell, et.al. [7] using an empirical formula by Vesic [8]. The mat is designed so that it can be rolled into a bundle for shipping (Figure 1). The site preparation requires removal of the first one-foot layer of the lunar surface soil.

Radiation Shielding

The radiation shielding is essentially made of flat space trusses similar to the floor trusses inside the habitat,

filled with lunar regolith, and each is wrapped with fabric to contain the regolith (Figure 1).

Materials

Aluminum Lithium 8090-T8771 is proposed for the interior floor trusses for its superior strength-to-weight ratio. Its relevant properties at room temperature are: [4,5]

Ultimate Strength:	64 ksi (441.28 MPa)
Yield Strength:	50 ksi (344.75 MPa)
Percent Elongation:	.5 to 2 %
Mass Density:	0.091 lb _m /in ³ (2,519 kg/m ³)
Modulus of Elasticity:	11.7x10 ⁶ psi (80,672 MPa)
Thermal Expansion Coeff.:	13x10 ⁻⁶ /°F (23x10 ⁻⁶ /°C)

Magnesium alloy ZCM 711, high in its strength-to-weight ratio, is selected for the exterior support structure and the foundation mat. This material has a relatively low combustion temperature and therefore is unsafe to use inside the habitat where oxygen is present. The relevant properties of this material are: [4,5]

Ultimate Strength:	40 ksi (275 MPa)
Yield Strength:	27 ksi (185 MPa)
Percent Elongation:	12 %
Mass Density:	0.065 lb _m /in ³ (1,795 kg/m ³)
Modulus of Elasticity:	6,500 ksi (45 GPa)
Thermal Expansion Coeff.:	15x10 ⁻⁶ /°F (27x10 ⁻⁶ /°C)

Structural Dead Weight

The following weight estimate is based on a very basic conceptual design and therefore gives only a rough idea of the amount of materials to be shipped to the moon:

Upper Floor Truss:	1,320 lb _m (600 kg) of Al-Li
Lower Floor Truss:	950 lb _m (430 kg) of Al-Li
Support Structure:	560 lb _m (254 kg) of ZCM
Foundation Mat:	4,400 lb _m (1,995 kg) of ZCM _____
Total	7,230 lb _m (3,280 kg)

Field Erection

The following is a general outline of the field erection procedure with much further refinement yet to be developed.

1. A flat level surface is prepared at the site after the removal of the first one-foot layer of the lunar soil.
2. The top horizontal truss for the radiation shielding (roof) is fully stretched and is resting on the vertical

- trusses (walls) which are not yet fully stretched.
3. The top horizontal truss is raised to its final elevation by jacking with the vertical trusses (walls) fully stretched to their final configuration simultaneously.
 4. All fabric bags over the trusses were pre-attached and are now in their fully stretched shapes.
 5. All trusses are filled with lunar regolith from top by a bucket elevator.
 6. The foundation mat is laid.
 7. The habitat module is placed on the mat.
 8. The cylindrical air enclosure is inflated to its final volume while the pre-attached interior floors and the exterior support frame are also fully stretched to their final configuration.

Discussion

Aluminum-Lithium Al-Li 8090 is selected for the interior floor trusses for its superior strength-to-weight ratio to minimize shipping weight. The strength-to-weight ratio of Magnesium ZCM 711 is higher than that of Al-Li 8090 but ZCM 711 is unsafe to use inside the habitat due to its low combustion temperature. ZCM 711 is therefore selected for only the exterior support frame and the mat. Most Magnesium alloys including ZCM 711 are known to be weak in corrosion resistance. This presents no problem on the moon for there is no oxygen nor moisture outside the habitat.

While the Aluminum alloy is known to have superior resistance to embrittlement at low temperature [11] the Magnesium alloy ZCM 711 loses its ductility at low temperature and is therefore unsuited for load carrying in tension. It is therefore chosen for the support structure and the mat only where materials are primarily under compression.

The thickness of the first layer of the lunar surface soil to be removed was arbitrary in the design. Generally speaking, the soil bearing capacity increases with increase in depth but at the expense of increased site preparation.

THE GUYED TOWER

General Concept

The general concept of the proposed lunar guyed tower is essentially the same as that of typical guyed towers found on Earth (Figure 2). Unlike the design of tower structures on Earth, a lunar guyed tower requires no resistance to lateral wind loads for there is no wind on the moon. The only significant environmental lateral loading is perhaps due to a

moonquake. It is possible, however, that significant stress and deformation occur due to severe daily (lunar) temperature variations. A structural analysis under this thermal loading (with the gravity loading properly combined) should be performed after the structure is better defined and a cyclic heat transfer and temperature-time history throughout the structure is assessed. In an effort to mitigate this temperature effect two different materials were selected, i.e., Magnesium alloy for the mast and Aluminum-Lithium for the guys. This temperature effect is further discussed in a later part of the report.

Design Requirements

Given below is a list of a very preliminary requirements/constraints for the design of the lunar guyed tower.

1. A payload of 1,100 lb_m (500 kg) at approximately 2,000 feet above ground (given by M.L. Roberts, Systems Engineering Division, JSC).
2. A minimized total structural dead weight.
3. Satisfaction of the Specification for the Design, Fabrication, and Erection of Structural Steel for Buildings (AISC) [6].
4. Efficient for shipping and field erection.
5. Resistant to moonquake.
6. Resistant to daily (lunar) temperature variations.
7. An elevator to the top of the tower.

The Mast

The mast is composed of identical 9-foot truss sections made of standard tubular members of Magnesium alloy C2M 711. The three cords and the braces make the cross sectional shape of the mast an equilateral triangle (Figure 2). The bottom 9-foot truss section is different from the rest (Figure 3) where larger openings are provided to facilitate traffic to the interior of the mast.

The Guys

The use of wire ropes made of Aluminum Lithium 8090 is proposed for the guy cables. A simple calculation (assuming uniform temperature variation throughout the mast and the guys) shows that, for a guy cable making an angle of approximately 58 degrees with horizontal and having its thermal expansion coefficient 23/27 of that of the mast [5], stresses induced in the guys and the mast are minimal due to the diurnal temperature variations. Using the formulas provided in [16] the maximum and minimum temperatures are approximately 240°F (388°K) and -248°F (118°K) respectively

(depending also on the location on the lunar surface). These temperature extremes corresponds to a temperature range of 487°F (270°K) [16]. Since the metal tower is more reflective than the lunar surface the above temperature variation of the lunar surface can be used as the upper and lower bounds of temperature variation of the tower structure.

Since there is no lateral wind load on the lunar surface vertical locations of upper guy connections are determined solely by the buckling consideration of the mast. Thus the vertical spacings for the upper guy connections are determined by the Euler buckling formula [9] with a factor of safety of 2, i.e., $L = n (EI/2P)^{1/2}$ where P is conservatively taken as the axial compression at the lower end of the unbraced length L (the structural dead weight makes the axial compression at the lower end greater than at the upper end). To include the effect of the mast's own body weight on its critical buckling load a numerical procedure was formulated but is not presented in this report for lack of verification. Published information on column buckling including the column's own body weight can not be found at time of this writing [15]. This may be attributed to its relatively small effect in design (on Earth) as compared to that due to wind load.

The guys are to be sized to provide a lateral support capacity at their upper connections equal to 10% of the axial compression in the mast (to be considered fully effective). According to a recommended practice [10], the guys are to be pretensioned to 1/8 of their breaking strength. These pretensions in guys contribute to the axial compression in the mast (Figure 2).

Foundation

An equilateral triangular mat is sized to provide a bearing area at the base of the mast with a factor of safety of 4 against the ultimate bearing capacity of the lunar soil at approximately one foot below grade [7,8]. Immediately above the mat is a ball joint to effect a hinged support so that there will be no over-turning moment to be resisted by the foundation during a moonquake [14] (Figure 3).

All anchors of the guys should provide an ultimate up-lift capacity based on the breaking strength of the guys. A safety device is to be installed in series at the lower connection of each guy. The device should disengage when tension in the guy reaches a certain fraction of the guy's breaking strength. Experience in up-lift and lateral load capacities of piled or coil anchors on lunar surface and their installation is not available at time of this design study. Estimates of these capacities can be made using some existing

numerical methods generally available within today's civil engineering practice. A less efficient alternative for the anchors would be the use of the so-called "deadman" concrete blocks. This anchoring alternative may become more attractive when cast-in-place lunar concrete becomes feasible.

Materials

The same Aluminum Lithium 8090-T8771 and Magnesium alloy ZCM 711 described earlier in this report for the lunar habitat are proposed for the guys and the mast respectively.

Structural Dead Weight

The total structural mass is estimated at about 9,000 lb_m with the following breakdowns: (excluding payload and anchors)

Mast:	8,620	lb _m (3,910 kg)	of ZCM
Guys:	250	lb _m (113 kg)	of Al-Li
Pedestal:	160	lb _m (73 kg)	of ZCM
Total:	<u>9,030</u>	lb _m (4,096 kg)	

Field Erection

The following field erection scheme for the guyed tower is proposed (Figure 4).

1. Prepare a flat level surface at the site after removal of the first one-foot layer of the lunar soil.
2. Set the foundation mat and the hinged-support assemblage (pedestal) on the prepared surface.
3. Position the top 9-foot section of the mast onto the pedestal.
4. Connect the three top-level guy cables to the top 9-foot section of the mast and to their anchors with some reels with a controlled tensioning device.
5. Install the hydraulic jack inside the mast.
6. Jack up the top 9-foot section of the mast until its bottom is clear for the next 9-foot section to be assembled around the hydraulic jack immediately below.
7. Assemble the next 9-foot section.
8. Repeat steps 6 and 7 and connect guy cables as designed until the last section.
9. Assemble the last bottom 9-foot section and secure to the support pedestal.
10. Remove the hydraulic jack.
11. Install the elevator.
12. Remove the three temporary members which were holding the support pedestal to the foundation mat.
13. Adjust tension in guys for mast's verticality.

Discussion

Due to the loss of ductility of the Magnesium alloy ZCM 711 at low temperature, ZCM 711 is selected only for the mast since the mast is primarily under compression. Aluminum-Lithium is selected for the guys (in tension) for their known ductility, modulus of elasticity, and strength preservation at both high and low temperatures [11,13]. Possibility of creeping of the mast and guys at the upper temperature extreme should be studied.

Numerical predictions of capacity and driveability of piled anchors are well developed for design on Earth. Applicability of these numerical techniques to lunar surfaces needs further investigation.

Dynamic amplification and response of the tower to moonquake should be analyzed and the structural integrity be verified at the two temperature extremes [14,16].

The in-situ guy cables are pretensioned to 1/8 of the breaking strength of the cable as generally practiced for guyed towers on Earth [10]. This pretensioning should be studied and possibly revised for guy cables on the moon.

Each 9-foot section of the mast can be made collapsible for shipping, stretched and assembled around the hydraulic jack at the site.

The daily (lunar) temperature variations of the structure can be reduced to a certain extent by coating the structure with some reflective material. The reflective coating reduces the net solar heat influx and thus lowers the upper extreme of the temperature variation.

It may be necessary to avoid direct contact between the mast (of Magnesium alloy) and the guys (of Aluminum-Lithium) at the upper guy connections for possible galvanic corrosion due to the two dissimilar metals.

A constant angle (approx. 60°) from the horizontal is proposed for all the guys to mitigate the stresses due to the temperature variations (in contrast to a single anchor for each vertical plane of guys, Figure 2).

REFERENCES

1. Yin, P.K., "A Preliminary Design of Interior Structure and Foundation of an Inflatable Lunar Habitat," Final Report, NASA/ASEE Summer Faculty Fellowship Program, Johnson Space Center, 1989.
2. Yin, P.K., "Structural Design of an Inflatable Lunar Habitat," Engineering, Construction, and Operations in Space II, Proceedings of Space 90, ASCE, 1990.
3. Structural Dynamics Research Corporation (SDRC), User's Guide, I-DEAS Level 4, Supertab .Engineering Analysis, 1988.
4. Charles, J.A., Crane, F.A.A., "Candidate Materials for Aircraft Structures," Selection and Use of Engineering Materials, 2nd Ed., Section 15.4, Butterworths, 1989.
5. Gere, J.M., Timoshenko, S.P., "Mechanical Properties of Materials," Mechanics of Materials, 2nd Ed., Appendix H, Brooks/Cole Engineering, 1984.
6. American Institute of Steel Construction, Inc. (AISC), Manual of Steel Construction, 8th Edition, 1980.
7. Mitchell, J.K. et.al., "Mechanical Properties of Lunar Soil: Density, Porosity, Cohesion, and Angle of Internal Friction," Proceedings of the Third Lunar Science Conference, Vol. 3, pp 3235-3253, The M.I.T. Press, 1972.
8. Vesic, A.S., "Bearing Capacity of Shallow Foundations," Foundation Engineering Handbook, Van Nostrand Reinhold Co., 1975.
9. Gere, J.M., Timoshenko, S.P., "Chapter 11: Columns," Mechanics of Materials, 2nd Ed., Brooks/Cole Engineering, 1984.
10. Zar, Max, "Section 24: Towers," Structural Engineering Handbook, McGraw-Hill Book Company, 1968.
11. Glazer, J., et al, "Cryogenic Mechanical Properties of Al-Cu-Li-Zr Alloy 2090," Presented at the International Cryogenic Materials Conference, Cambridge, MA, August 12-16, 1985.
12. Wesselski, C.J., "Expandable Pallet for Space Station Interface Attachments," U.S. Patent number 4,805,368, Feb. 21, 1989.
13. American Society of Metals (ASM), Metals Handbook Desk Edition, 1985.
14. McCaffrey, R.J., Hartmann, A.J., "Dynamics of Guyed Towers," J. of the Structural Division, Proceedings of the ASCE, Vol. 98, No. ST6, June, 1972.
15. Goldberg, J.E., Gaunt, J.T., "Stability of Guyed Towers," J. of Structural Division, Proceedings of the ASCE, Vol. 99, No. ST4, April, 1973.
16. Binder, Allen, Lunar Data Base, Lockheed Engineering & Sciences Co. Inc., April, 1990.

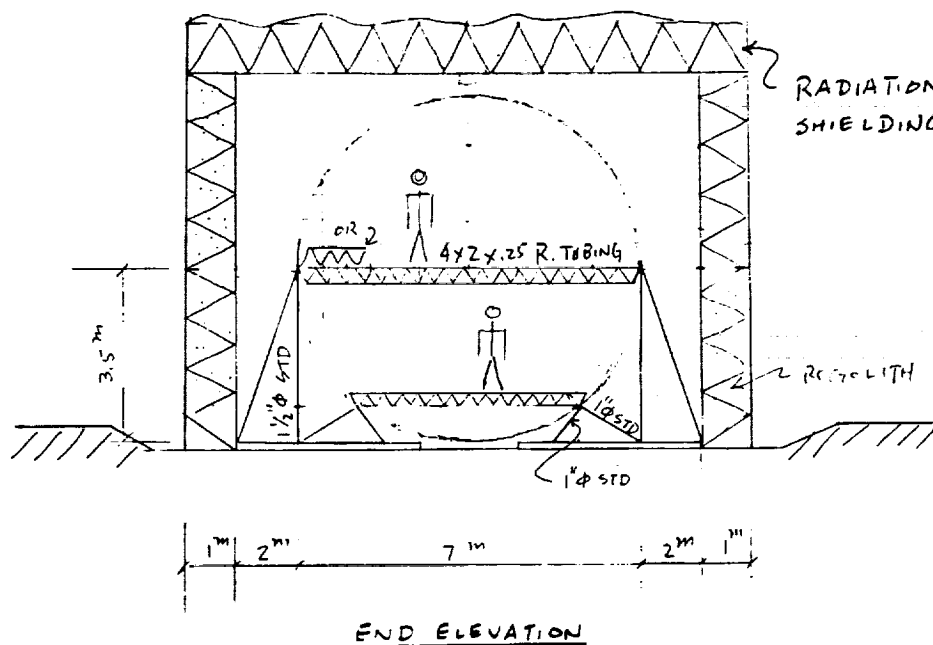
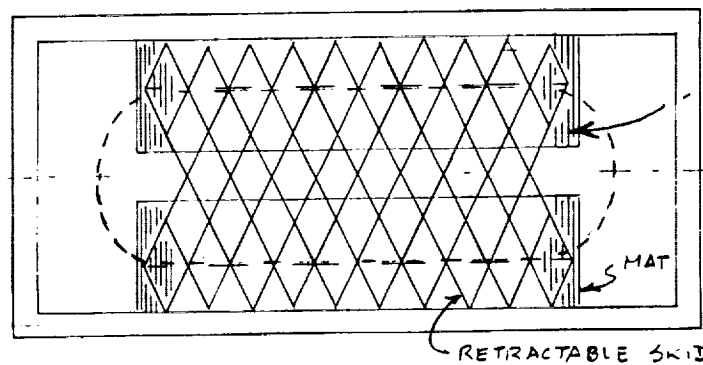
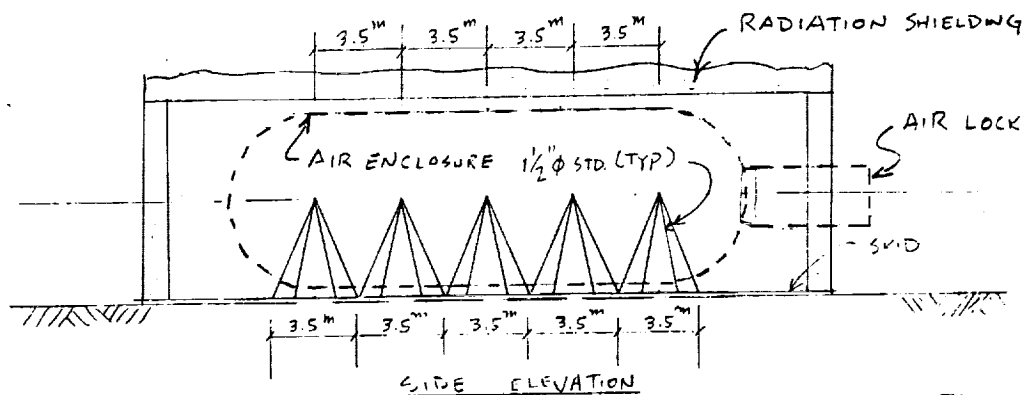


Figure 1. The horizontal cylindrical lunar habitat.

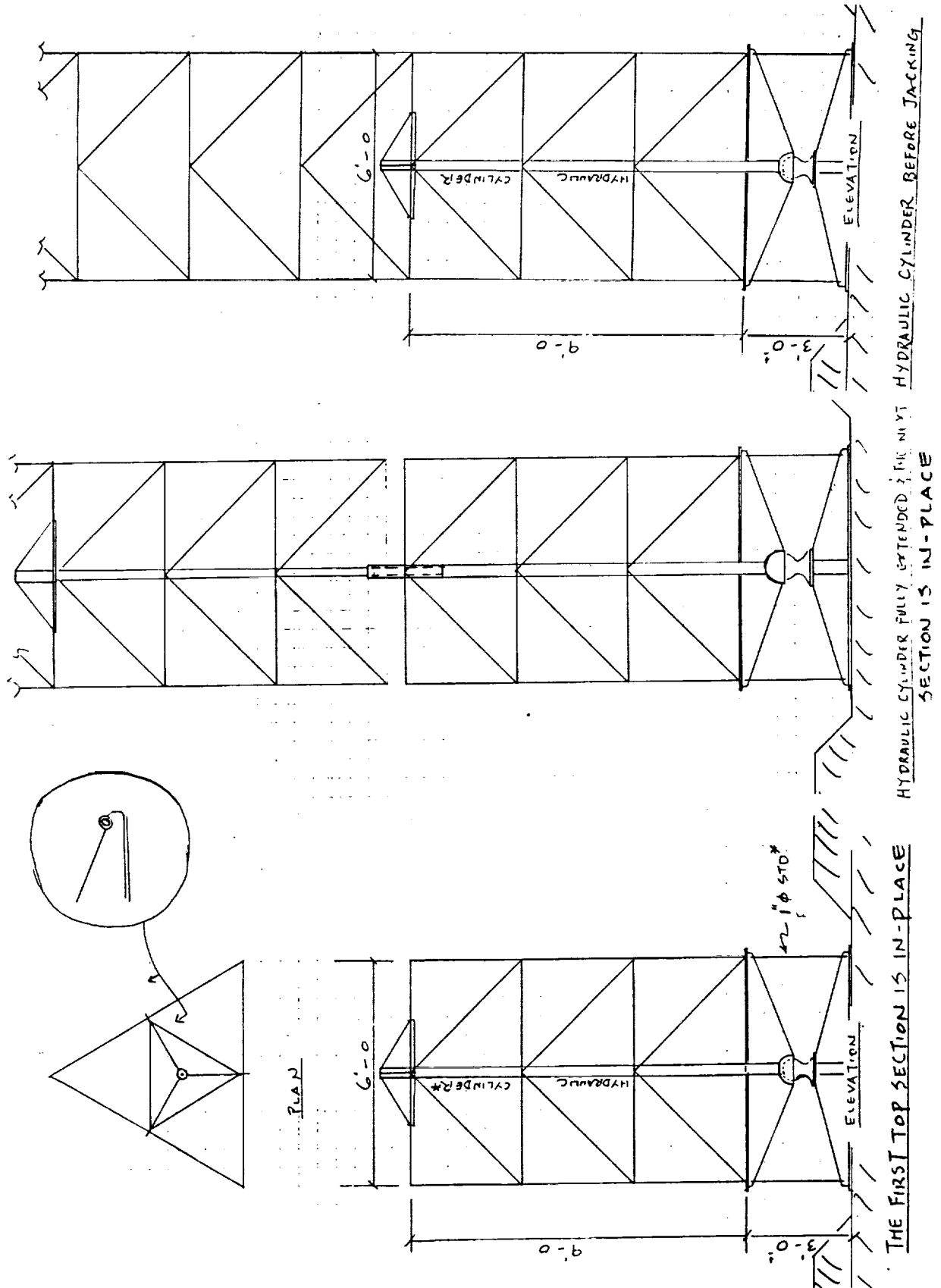


Figure 4. The guyed tower's installation sequence.



National Aeronautics and
Space Administration

REPORT DOCUMENTATION PAGE

1. Report No. NASA CR 185637		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle NASA/ASEE Summer Faculty Fellowship Program--1990 Volume 2				5. Report Date December 1990	
				6. Performing Organization Code	
7. Author(s) Richard B. Bannerot and Stanley H. Goldstein, Editors				8. Performing Organization Report No.	
				10. Work Unit No.	
9. Performing Organization Name and Address University of Houston-University Park Houston, TX 77204				11. Contract or Grant No. NGT 44-005-803	
				13. Type of Report and Period Covered Contractor Report	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546				14. Sponsoring Agency Code	
				15. Supplementary Notes	
16. Abstract The 1990 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by the University of Houston-University Park and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are (1) to further the professional knowledge of qualified engineering and science faculty members; (2) to stimulate an exchange of ideas between participants and NASA; (3) to enrich and refresh the research and teaching activities of participants' institutions; (4) to contribute to the research objective of the NASA Centers. Each faculty fellow spent at least 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects performed by the faculty fellow during the summer of 1990. Volume 1 contains reports 1 through 14, and Volume 2 contains reports 15 through 28.					
17. Key Words (Suggested by Author(s))			18. Distribution Statement Unclassified - Unlimited		
19. Security Classification (of this report) Unclassified		20. Security Classification (of this page) Unclassified		21. No. of pages	
				22. Price	

For sale by the National Technical Information Service, Springfield, VA 22161-2171